Strong Gravitational Lensing as a Probe of Cosmology and Galaxy Evolution at Sub-Galactic Scales

by

Bingxiao Xu

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

September, 2017

Abstract

The main aim of my thesis work is to understand galaxy evolution at higher spatial resolution awarded by strong lensing. To achieve that, I have performed systematic studies focusing on three different aspects of the strongly lensed galaxies (giant arcs) in CLASH survey. In the first, I study the abundance of the strongly lensed galaxies in CLASH survey. I have devised an automated arcfinder to search giant arcs in CLASH images. After correcting the incompleteness and false positive rate, I find that at least for a sample of 20 clusters, the predicted arc abundance based on ΛCDM is remarkably consistent with the observed abundance, which is $\sim 4 \pm 1$. I also have performed extensive simulations to test the sensitivity of arc abundance to different physical parameters. I find that the arc abundance is more sensitive to the inner mass profile of galaxy clusters than the redshift distribution of background galaxies. The second part is to study the physical properties of the lensed galaxies including the stellar mass, star formation rate (SFR), specific star formation rate (sSFR), color, etc. The strong lensing generally selects galaxies that have

ABSTRACT

lower mass than those from field survey. Majority of CLASH lensed galaxies have mild star formation rate with median value $\sim 4\mathrm{M}_{\odot}\mathrm{yr}^{-1}$. I reconstruct the images of CLASH giant arcs in their source plane based on CLASH mass models and study the morphology of the galaxies in the source plane. Most of them lie in the irregular region in the Gini-M20 plane, which is consistent with that the low-mass galaxies at high-z typically have clumpy and irregular morphology. In the third study, I explore the sub-galactic structures (clumps) identified in the source plane reconstructed images. I study various physical properties of the sub-galactic structures (clumps) such as stellar mass, SFR, sSFR, color, size and compactness. I also have quantified the rest-frame UV luminosity function (LF) and stellar mass function (SMF) of clumps at different redshifts. The main results are: the clumps at high-z are typically more massive and compact; the high-z clumps have very similar structural properties to the dwarf spheroidals at low-z, which implies possible evolutionary link between two species; the faint end and low mass end slopes of rest-frame UV LF and SMF are flatter than those of galaxy at similar redshifts; the high mass end of clumps SMF at high-z is very well consistent with the SMF of bulge in late-type disk galaxies with similar abundance, which suggests the massive clumps at high-z are likely progenitors of today's bulge; intriguingly, the most massive clumps at high-z are also rest-frame UV brightest, which implies that the mass assembly of bulges proceeds mainly via some dissipative process

ABSTRACT

rather than the “dry” growth which is predicted by the some theoretical models; the spatial distribution of clumps and its evolution are consistent with that the clumps do not migrate towards the galactocenter or migrate very slowly in low-mass galaxies. Combining the above evidence, I propose a brand new picture of galaxy evolution to simultaneously explain the formation of the today’s Hubble sequence, formation of the galactic bulges and origin of the dwarf spheroids in dense environments.

Primary Reader and Advisor: Dr. Marc Postman

Secondary Reader: Prof. Colin Norman

Acknowledgments

I would evaluate the past years in Baltimore as the most significant time in my life so far. Not only do I luckily have the opportunity to work together with the superb scientists in most interesting and exciting fields in astrophysics, but also my skill, vision, perseverance or even personality gets largely forged by different people and experience. I am grateful to all the people who have offered me assistance in academia and daily life in my PhD life.

I would first express my gratitude to my dear parents. It would have been impossible for me to make it through without your enduring support, care and encouragement. Especially for my mom, your encouragements are always my source of strength to rise to the occasion.

To my advisor Dr. Marc Postman, thank you for guiding me into the exciting arc statistics and high-z substructure fields; thank you for teaching me not only the knowledge and skill, but also some unique characteristics to be a good researcher that are impossible to learn from anywhere else, and I will definitely benefit from the way of thinking you have taught me in the rest of my life;

thank you for directly providing me insight in my research. My most important part of work on sub-galactic structure would have not been possible without your perceptive suggestion on LF and SMF. Thank you for all of these. It was really my pleasure to work with you in the past five years!

To my co-advisor Prof. Colin Norman, thank you for teaching me cosmology and dark matter physics, which was still very influential on my understanding of structure formation and galaxy evolution; thank you for not giving me up when I was out of form, rather, offering me precious opportunity to join CLASH; thank you for pushing me forwards when I was not in the right track; Thank you for all of these, I would really appreciate it!

I am grateful to Kevin Fogarty. Thank you for teaching me to set up and use iSEDfit. Thank you for generously allowing me to log on your server to run SED fitting. Thank you for helping me run the codes on your computer at my desperate urgent time. Thank you very much and I would really appreciate it!

I am grateful to all my collaborators Massimo Meneghetti, Carlo Giocoli, Adi Zitrin, Stella Seitz, Wei Zheng, Keiichi Umetsu, etc. Thank you for your invaluable contributions to my work and thoughtful suggestion and feedback.

I am grateful to my thesis committee members: Prof. Brice Menard, Dr. Jennifer Lotz and Dr. Dan Coe. Thank you for your precious time spending on reading my manuscript and offering me insightful suggestions to my thesis works.

ACKNOWLEDGMENTS

I am grateful to my dear fellow friends. Thank you for accompanying me through those happy and hard times, leaving me several memorable silhouettes.

Contents

List of Tables

List of Figures

LIST OF FIGURES

Chapter 1

Introduction

In the last decades observations have shown that our universe consists of only $\sim$ 5% of the baryonic matter that we can directly observe with instruments. The rest of its content is associated with non-visible matter, including $\sim$ 22% dark matter and $\sim$ 73% dark energy (e.g. see Komastu et al. (2011)). The existence of dark matter was probed through the rotation curves of galaxies, which are flat in the outskirts, and implying the need of extra source of gravitational potential aside from the observed baryonic ones (e.g. see Rubin & Ford (2011)). The existence of dark matter was also a necessity to to reproduce the observed large scale structure (LSS) of the Universe in N-body numerical simulations (Springel, 2005).

Galaxies, as the basic building block of structure formation, are forming inside the skeletons of dark matter halos as time elapses. Galaxies are composed

of dark matter and baryonic matter such as stars and gas, with typical stellar masses between $10^8 - 10^{12} M_\odot$. Hubble (Hubble, 1936) proposed a morphological classification of nearby galaxies based on their observed morphology, in which galaxies are divided in four main classes: Ellipticals (E); Lenticulars (S0); Spiral galaxies (S) and Spiral barred (SB); Irregular galaxies (Irr). (e.g. Binney & Merrifield (1998)). This is known as Hubble sequence or Hubble tuning fork shown in Figure 1.1. It was thought to be representative of the evolutionary sequence of the galaxies, so the ellipticals, lying on the left side of this sequence were called early type galaxies (ETGs), while those on the right side were referred to as late type galaxies (LTGs). Now this interpretation of the Hubble sequence is completely discredited (Baldry, 2008). For galaxies with settled regular morphology, there are two typical components: disks and spheroidals. The massive elliptical galaxies are spheroidals themselves; whereas the disk galaxies host spheroidal components in their center called galactic bulges. The stellar populations in spheroidal components are typically much older than those in disk components, with little current star formation activity and lower metallicity, which implies that they form in earlier stage than those in the disks. When and how do these spheroidals form? In classic picture of galaxy formation, they were proposed to form in galaxy mergers (Milos & Hernquist, 1996; Hopkins et al., 2012), which is however disfavored by growing observational evidence. More feasible mechanisms are yet to come

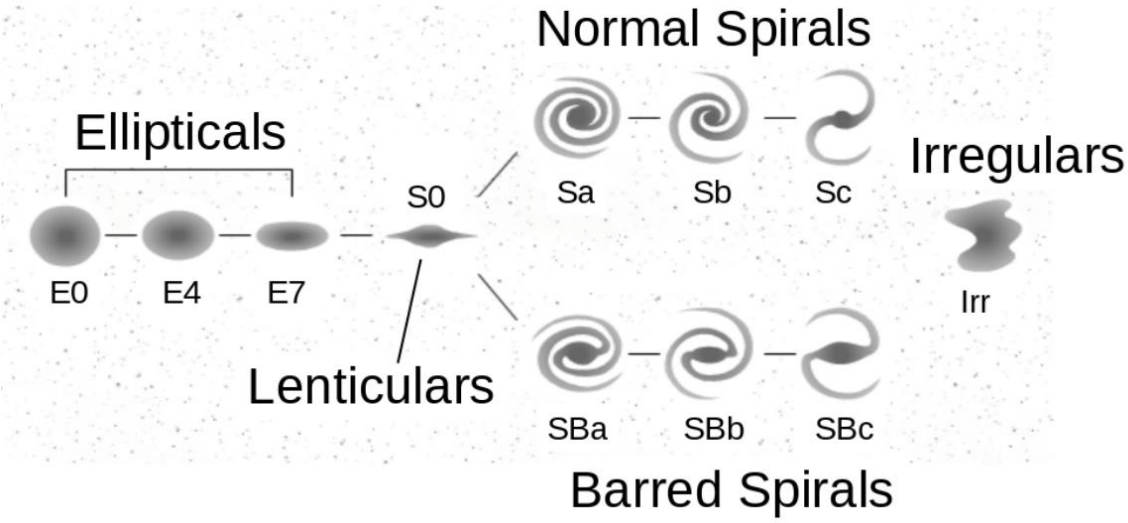


Figure 1.1: Diagram of the Hubble sequence. Figure from https:// concept-draw.com/a2353c3/preview.

into being.

In the early Universe, young galaxies appear to look very different from today's galaxies: they typically show irregular and clumpy morphology (Elmegreen et al., 2005; Guo et al., 2015). How the irregular and clumpy morphology settles down to today's regular morphology is still a mystery. More intriguingly, the local galactic bulges typically contain stellar mass of $10^8 - 10^{10}\mathrm{M}_\odot$, and have size of $\sim 1\mathrm{kpc}$. The study of high-z clumpy galaxies have revealed that the typical mass and size of the substructures in the clumpy galaxies are comparable to those of bulges (Elmegreen et al., 2009; Forster Schreiber et al., 2009; Guo et al., 2012). Is this just a coincidence? Or is it related to the unknown process of settling the modern Hubble sequence? To understand all the above open questions, we need detailed information of the sub-galactic structures at high-z

and its evolution. However, the spatial resolution becomes the main obstacle posed in front of the studies of high-z sub-galactic structures. Even with Hubble Space Telescope (HST), resolving sub-kpc substructures above $z_s > 2$ is a mission impossible due to the PSF blurring effect.

On the other hand, clusters of galaxies as the most massive gravitationally bound structures, are dark matter dominant systems. The huge mass budget in cluster of galaxies bends the light rays from galaxies behind and makes clusters of galaxies act as gravitational lenses. The gravitational lensing effect magnifies the flux and surface area of the background galaxies and allows us to probe galaxies at very high redshift and study the substructures in galaxies at high redshift at an unprecedented spatial resolution.

In this thesis, I present my research on galaxy evolution based on strong lensing effect, aiming to address the above open questions. The thesis is organized as follows. In Chapter 1, I describe the basic theory and mathematics of gravitational lensing. In Chapter 2, I describe my work on studying cluster lensing efficiency and its dependence on various physical effects, aiming to investigate how the dark matter distribution in clusters affects the abundance of large lensed galaxies and the compatibility of the abundance of large lensed galaxies between the simulations based on ΛCDM and the observations. In Chapter 3, I describe the integrated properties and morphology of the lensed galaxies at high-z, to understand what the morphology of the high-z galaxies

are at high spatial resolution. In Chapter 4, I describe the physical properties of the substructures in galaxies at high-z, including their stellar mass, star formation rate, specific star formation rate, multiplicity, rest-frame UV luminosity function and stellar mass function, to address what the distributions of physical properties of sub-galactic clumps are as a function of redshift. In Chapter 5, I describe the radial distribution and radial variation of physical properties of clumps to address what the spatial distribution of clump properties can tell us about their role in galaxy evolution. Overall discussions and a summary of my findings are presented in Chapter 6 & 7.

1.1 The Principles of Gravitational Lensing

Gravitational lensing theory provides a complete analytic description of lensing phenomena at the first order with a few simplifying hypotheses (e.g. Schneider (1992); Narayan & Bartelmann (1996)).

First of all, it assumes that a **weak stationary gravitational field** is associated with the lens, which means that the Newtonian potential of the lens is weak ($\phi_{lens} \ll c^2$) and the relative velocities of the lensing systems (i.e. source, lens and observer) are small with respect to the speed of light ($v_S, v_L, v_O \ll c$) (Narayan & Bartelmann, 1996).

Then it assumes that the physical size of the lens is much smaller than the distances between lens, source and observer. This assumption is called **thin lens approximation**.

These assumptions always hold in astrophysical lensing systems. For instance, if a source at $z_s \sim 1$ lensed by a galaxy cluster at $z_{cl} \sim 0.3$, the distances between source, lens and observer are of the order of Gpc, which is much larger than the typical sizes of clusters of order of Mpc.

Figure 1.2 presents a typical lensing diagram, with a lens placed at the angular-diameter distance D_d and a source at the distance D_s. According to the thin lens approximation, we can write

$$\vec{\beta} D_s = \vec{\theta} D_s - \hat{\vec{\alpha}} D_{ds}, \tag{1.1}$$

where $\vec{\beta}$ is the angular position of the source on the sky and $\vec{\theta}$ is the apparent angular position of the source. This equation is called lens equation, which is a mapping equation from the image plane to the source plane. The deflection angle $\hat{\vec{\alpha}}$ contains all the information of the mass distribution of the lens and is given by (see Narayan & Bartelmann (1996)):

$$\hat{\vec{\alpha}}(\theta) = \frac{2}{c^2} \int \vec{\nabla}_\perp \Psi dz \tag{1.2}$$

where Ψ is the gravitational potential of the lens and $\nabla_\perp \Psi$ is the gradient of Ψ

perpendicular to the light path.

Using the thin lens approximation, we describe the mass of the lens through its surface density Σ, which is the 2D projected lens density $\rho(\vec{r})$ on the image plane (see Narayan & Bartelmann (1996)):

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, z) dz \tag{1.3}$$

where $\xi = D_d \theta$.

Then the deflection angle is given by:

$$\hat{\vec{\alpha}}(\vec{\xi}) = \frac{4G}{c^2} \int \frac{(\vec{\xi} - \vec{\xi}')\Sigma(\vec{\xi}')}{|\vec{\xi} - \vec{\xi}'|^2} d^2\xi' \tag{1.4}$$

We can introduce the critical surface density Σ_{cr} as (see Narayan & Bartelmann (1996))

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_s}{D_{ds} D_d} \tag{1.5}$$

and then the deflection angle for a symmetric circular lens with constant Σ becomes

$$\hat{\alpha}(\theta) = \frac{\Sigma}{\Sigma_{cr}} \theta. \tag{1.6}$$

If the lens has $\Sigma > \Sigma_{cr}$ then $\alpha > \theta$, and the source is *strongly* deflected. In this case we are in the *Strong Lensing* regime, where highly deformed images and multiple images of the same source are generated. If $\Sigma \ll \Sigma_{cr}$ then $\alpha < \theta$,

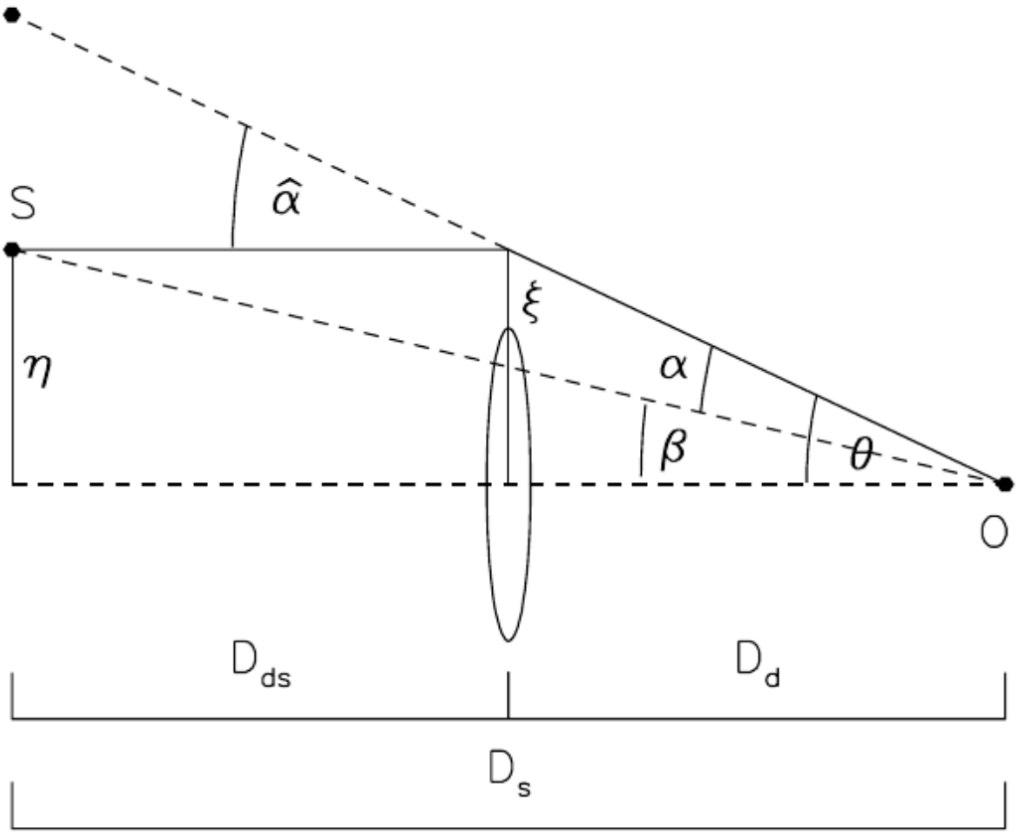


Figure 1.2: Diagram of a gravitational lensing system. Due to the gravitational lensing deflection, the observer sees the source in the position I on the lens plane, while its actual position is S, on the source plane. Due to the thin lens approximation we can assume that the deflection takes place in correspondence of the lens plane, and that the light emitted by the source travel undisturbed before and after this deflection. The dashed line is the optical axis of the system, and is perpendicular to the lens and source planes, passing through the observer position. The angular positions on the source and lens planes are measured relative to this axis. Figure from Narayan & Bartelmann (1996).

and we are in the *Weak Lensing* domain, where the the image shape is weakly distorted by the lens.

Then we introduce the effective lensing potential $\psi(\vec{\theta})$ (see Narayan & Bartelmann (1996)), defined as a rescaled projection of the three-dimensional Newtonian potential on the lens plane

$$\psi(\vec{\theta}) = \frac{D_{ds}}{D_s D_d}\frac{2}{c^2}\int \Psi(D_d\vec{\theta}, z)dz \tag{1.7}$$

The gradient of ψ along θ is equal to the reduced deflection angle:

$$\vec{\nabla}_\theta\psi(\vec{\theta}) = \vec{\nabla}_\perp\frac{D_{ds}}{D_s}\frac{2}{c^2}\int \Psi(D_d\vec{\theta}, z)dz = \tag{1.8}$$

$$= \frac{D_{ds}}{D_s}\frac{2}{c^2}\int \vec{\nabla}_\perp\Psi(D_d\vec{\theta}, z)dz = \alpha(\vec{\theta}). \tag{1.9}$$

Thus

$$\vec{\nabla}_\theta\psi(\vec{\theta}) = \alpha(\vec{\theta}). \tag{1.10}$$

In addition, the Laplacian of ψ is directly proportional to the so-called convergence function $\kappa(\vec{\theta})$, defined as the ratio between the surface density $\Sigma(\vec{\theta})$ and the critical surface density Σ_{cr}

$$\vec{\nabla}^2_\theta\psi = 2\frac{\Sigma(\vec{\theta})}{\Sigma_{cr}} = 2\kappa(\vec{\theta}) \tag{1.11}$$

Strong lensing occurs where the surface mass density $\Sigma(\vec{\theta})$ is larger than

the critical surface density Σ_{cr} , thus the convergence $\kappa(\vec{\theta}) > 1$: this condition marks the *strong lensing* domain, while $\kappa(\vec{\theta}) \ll 1$ marks the region of *weak lensing* regime. As previously stated, the lens equation is a *mapping equation* which maps each element $\delta\vec{\theta}$ of the image plane into a element $\delta\beta(\vec{\theta})$ on the source plane as follows

$$\delta\vec{\beta} = \frac{\partial\vec{\beta}}{\partial\vec{\theta}}\delta\vec{\theta} = \left(\delta_{ij} - \frac{\partial\alpha_i(\vec{\theta})}{\partial\theta_j}\right)\delta\vec{\theta} = \left(\delta_{ij} - \frac{\partial^2\psi(\vec{\theta})}{\partial\theta_i\partial\theta_j}\right)\delta\vec{\theta} = A\delta\vec{\theta} \tag{1.12}$$

The indexes $i, j = 1, 2$ are for the 2D components of the angles on the image and the source planes. The matrix A corresponds to the Jacobian matrix of the transformation (see Narayan & Bartelmann (1996)).

For simplicity we adopt the following shorthand notation for the second derivatives of the effective potential

$$\frac{\partial^2\psi(\vec{\theta})}{\partial\theta_i\partial\theta_j} = \Psi_{ij} \tag{1.13}$$

Using the Jacobian matrix A, we define the so-called Shear matrix as (see

Narayan & Bartelmann (1996))

$$\left(A - \frac{1}{2} trA \cdot I\right)_{ij} = \delta_{ij} - \Psi_{ij} - \frac{1}{2}(1 - \Psi_{11} + 1 - \Psi_{22})\delta_{ij} \tag{1.14}$$

$$= -\Psi_{ij} + \frac{1}{2}(\Psi_{11} + \Psi_{22})\delta_{ij} = \begin{pmatrix} -\frac{1}{2}(\Psi_{11} - \Psi_{22}) & -\Psi_{22} \\ -\Psi_{12} & \frac{1}{2}(\Psi_{11} - \Psi_{22}) \end{pmatrix} = - \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} \tag{1.15}$$

This matrix describes the asymmetric distortion of the shape of lensed images. $\gamma = (\gamma_1, \gamma_2)$ is the shear pseudo-vector. Its components are:

$$\gamma_1(\vec{\theta}) = \frac{1}{2}(\Psi_{11} - \Psi_{22}) \tag{1.16}$$

$$\gamma_2(\vec{\theta}) = \Psi_{11} = \Psi_{22} \tag{1.17}$$

while the eigenvalues of the shear matrix are

$$\pm\, \gamma = \pm\sqrt{\gamma_1^2 + \gamma_2^2} \tag{1.18}$$

The convergence describes the isotropic magnification of the lensed images relative to the source; whereas the shear describes the anisotropic deformation which stretches the shapes of the images along a particular direction, given by the phase of $\vec{\gamma}$.

The total magnification μ of the lensed images is given by

$$\mu = \frac{1}{detA} = \frac{1}{(1-\kappa)^2 - \gamma^2} \tag{1.19}$$

The family of points where the magnification is formally infinite on the image plane (i.e. where $detA(\vec{\theta}) = 0$) defines the so-called *critical lines*. λ_t, λ_r are defined as:

$$\lambda_r = \frac{1}{\mu_r} = 1 - \kappa + \gamma \tag{1.20}$$

$$\lambda_t = \frac{1}{\mu_t} = 1 - \kappa - \gamma \tag{1.21}$$

the critical lines are where $\lambda_t = 1/\mu_t = 0$ and by $\lambda_r = 1/\mu_r = 0$. $\lambda_t = 0$ and $\lambda_r = 0$ defines the tangential and radial critical lines respectively. Lensed images close to these lines are strongly distorted tangentially and perpendicularly to the curves respectively. Mapping the critical curves to the source plane through the lens equation, we then get the so-called *caustic lines*. A source which lies close to a tangential caustic line will be mapped close to a tangential critical line and showing a strong tangential distortion, and similarly a source lying close to a radial caustic will result in a radial lensed image.

1.2 The CLASH Survey

The work presented in this thesis has been performed using the Hubble Space Telescope (HST) data from the "Cluster Lensing And Supernovae with Hubble Survey (hereafter CLASH), see Postman et al. (2012). This survey aims to constrain in details the total mass distribution and concentration of clusters of galaxies through combined strong and weak lensing analyses, high resolution photometric data together with the spectroscopic measurements for 25 massive clusters.

1.2.1 CLASH Cluster Sample

The full CLASH cluster sample contains 25 massive galaxy clusters. 20 of these clusters were selected using X-ray archival data. Since one of the main goal of the survey is to investigate the mass distribution and concentration of galaxy clusters, no lensing information was used to select these 20 clusters to avoid any bias towards strong lensing. These 20 massive clusters, with $T_x \geq$ 5 keV, are dynamically relaxed, as revealed from their well defined surface brightness peaks with concentric isophotes in the X-ray images. The other 5 clusters were selected as gravitational lenses with an exceptional lensing strength, to magnify high redshift galaxies and thus to shed light on the early Universe. The 25 galaxy clusters span the redshift range $0.18 < z_{cl} < 0.90$ and

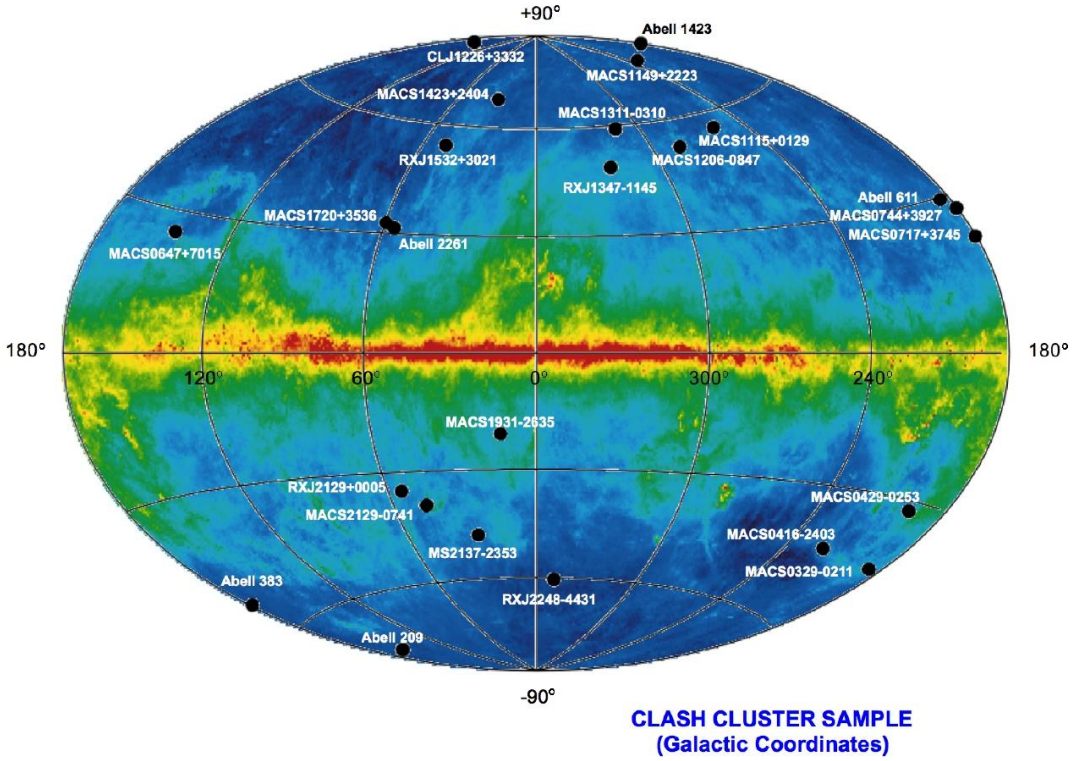


Figure 1.3: Distribution of the CLASH clusters on the sky map in Galactic coordinates (source: http://www.stsci.edu/ postman/CLASH/

mass range ~ 5 to $\sim 30 \times 10^{14} M_{\odot}$, thus allow us to probe the cluster concentration $c(M, z)$ over a wide range of mass and redshift. Figure 1.3 shows the position of the 25 CLASH clusters on the sky.

1.2.2 Observations

The CLASH survey is a 524-orbit Multi-Cycle-Treasury program (Postman et al., 2012), which has taken data over 3 annual HST observing periods, corresponding to cycles 18, 19 and 20. The clusters were observed in 16 broad-

band filters covering wide wavelength range of 2000-17000$\mathring{A}$ from the near-ultraviolet (NUV) to the near-infrared (NIR) using the Advanced Camera for Surveys (Ford et al. (2003), ACS) and the Wide Field Camera 3 (Kimble et al. (2008), WFC3) mounted on HST. The NUV range is covered with 4 broadband filters of the NUV channel of WFC3 (WFC3/UVIS), the optical range with 7 filters of the ACS/WFC and the NIR range is covered with 5 filters of the NIR channel WFC3/IR. Figure 1.4 shows the transmission curves of the 16 filters.

Each cluster was observed with 20 orbits, divided between the HST/ACS and HST/WFC3 filters. All the filters together cover a field of view (FOV) of $\sim$ 4.08 square arcminutes centered on the clusters. While the clusters were observed with one of the HST camera, parallel observations were taken with the other camera at $\sim$ 6 arcminutes distance from the cluster centre. The parallel field observations are used to search for the Type-Ia Supernovae out to $z \sim 2.5$. The observing times were set up to reach a 10σ depth of 26 AB mag in all the filters except for the F814W and the F160W filters in which the 10σ depth of $\sim$ 27 AB mag was required to identify high-z lensed candidate.

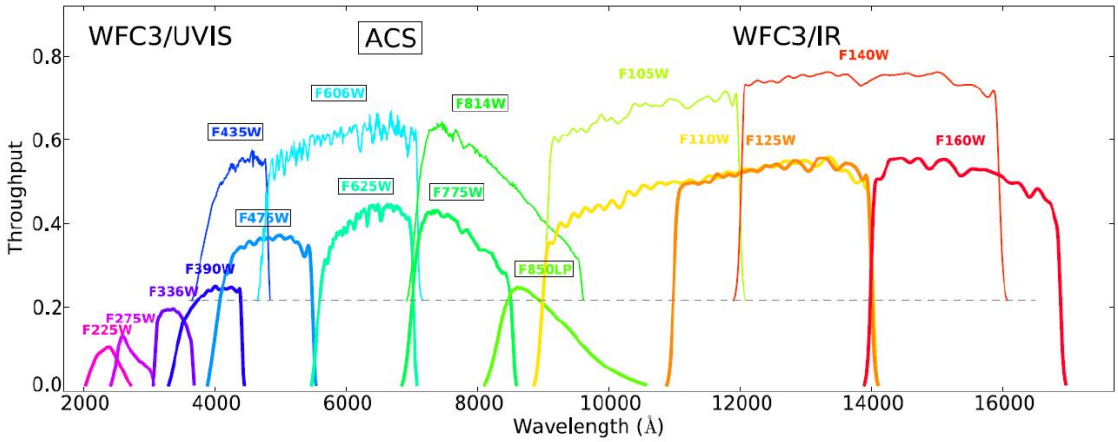


Figure 1.4: Throughput curves of the 16 HST/ACS and HST/WFC3 filters used in the CLASH survey (Postman et al., 2012).

Chapter 2

Arc Statistics

2.1 Introduction

The occurrence frequency of giant gravitationally lensed arcs – those most elongated, highly non-linear lensing features – is sensitive to the matter distribution within the cores of galaxy clusters. The statistics of giant arcs can thus provide useful tests of the structure formation. Cosmological models can be tested by comparing the observed giant arc abundance with the expected abundance from ray-tracing cosmological simulations. In an early study of arc statistics,

To see if the remaining discrepancies can be resolved, efforts need to be made on both theoretical and observational fronts. On the theoretical side, all effects which impact arc abundance should be included in the simulations

to make them more realistic. A straightforward example is to compare the lensing cross section in simulations with dark and baryonic matter against dark matter only simulations (Meneghetti et al., 2003; Puchwein et al., 2005; Killedar et al., 2012). On the observational side, larger, carefully selected cluster samples with ample redshift information are needed. All analyses will also benefit from the utilization of automated procedures for selecting giant arcs as rigorous comparisons must be done using an identical arc selection process for both the actual data and simulated data. In this respect, visual inspection, by which early arc statistics studies were conducted, is not an optimal approach as unquantifiable biases can potentially be introduced when classification is done by eye. Several groups have devised tools to search for arcs in an automated manner (Lenzen et al., 2004; Horesh et al., 2005; Alard, 2006; Seidel & Bartelmann, 2007). Most recently, Horesh et al. (2010, 2011) measured the observed abundance of arcs in a sample of 100 clusters observed with the Hubble Space Telescope (HST), using an automated and objective arc finder. The observed statistics were compared to those from a simulated dataset of cluster images. The simulated images were produced by ray-tracing through a large sample of clusters produced in N-body simulations, realistically simulating the observational effects, and then searching for arcs in these simulated clusters using the same arc-finding algorithm. Horesh et al. (2011) found excellent agreement between the observed and simulated arc statistics, particularly for their

main sub-sample of X-ray selected clusters at redshifts $0.3 < z < 0.6$. However, tension between the observations and simulations remained at other redshifts ranges, particularly for the subsample at $z < 0.3$. Moreover, none of the above groups has quantified the performance of their arc-finders, such as the arc detection completeness or the false positive rate. Without that information, the arc-finders' ability to predict the "true" arc abundance is limited.

In this paper, we measure the observed abundance of giant arcs from the CLASH (Cluster Lensing And Supernova survey with Hubble) sample (Postman et al., 2012). Giant arcs are found in the CLASH images, and in simulated images that mimic the CLASH data, using an efficient automated arc-finding algorithm whose selection function has been carefully quantified. CLASH is a 524-orbit multi-cycle treasury program that targeted 25 massive clusters with $0.18 < z < 0.90$. Twenty of the CLASH clusters are selected based on their X-ray characteristics. The X-ray selected CLASH sample contains clusters with $T_x \geq 5$ keV and with X-ray surface brightness profiles that have low asymmetry. The five remaining clusters were selected based on their expected lensing strength (large Einstein radii, typically $\theta_{Ein} > 30''$ for $z_s = 2$ or high magnification areas). Although the cluster sample is smaller than the one analyzed by Horesh et al. (2010, 2011), the CLASH observations are deeper, and photometric redshift information is available for all arcs brighter than about 26 AB mag (all the magnitudes hereafter are AB mag). In addition, our arc-finder is capa-

ble of detecting fainter arcs than previous studies. As a result, the total number of arcs that we find is comparable to that in the Horesh et al. studies. We simulate artificial clusters with the same mass and redshift range as the CLASH sample by using the N-body simulation-calibrated semi-analytic tool – MOKA (Giocoli et al., 2012), and directly from the high resolution, hydrodynamical simulations, MUSIC (Meneghetti et al., 2014), and perform ray-tracing simulation to prepare large sets of realizations for the simulated cluster images. We correct the raw arc counts in both the observations and simulations for incompleteness, false positive detections and arc elongation measurement bias. This allows us to conduct a direct comparison between the data and the simulations under different theoretical scenarios.

This paper is organized as follows: we describe the arc-finder algorithm and its implementation in section 2.2 and in the appendices; we demonstrate the arc-finder detection efficiency and overall performance in Section 2.3; we present the arc abundance results for the CLASH observations in Section 2.4; we describe the cluster simulation and ray-tracing calculations in Section 2.5; we compare the observed and simulated arc abundance results in Section 2.6, including specifically testing the dependence of the abundance on the source redshift distribution and $c - M$ relation in Section 2.7; a discussion and summary are given in Section 2.8 and 2.9, respectively. Throughout the paper, we adopt a ΛCDM cosmology with parameters $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.83$,

$H_0 = 100h \quad \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, and $h = 0.7$ (Planck Collaboration et al., 2014).

2.2 Conceptual Development of the arc-finder

In early works on arc statistics, arc detection was performed by visual inspection due, in part, to the complex shapes of arcs and the crowded environments in which they are found. An automated arc finding algorithm has three key advantages over visual search methods. First, the detection process is reproducible and can be implemented by anyone who learns how to run the code. Second, it can be applied to a large number of real and simulated images. Finally, the detection efficiency and false positive rate can be accurately quantified using artificial objects implanted in real data or using simulated images created by ray-tracing sources through lens models. The biggest challenge to developing such an algorithm is creating a definition of an arc for the purpose of detection that can be implemented in a robust manner using parameters that can be easily quantified from astronomical images.

An ideal arc finder should have the following characteristics:

1. The arc finder should be able to suppress image noise to enhance the contrast of real, low surface brightness arcs without significantly altering

the intrinsic shape characteristics of these faint objects.

2. The selection of pixels belonging to arcs should, if possible, not be based on a global fixed intensity threshold as the intensity can vary significantly across a lensed image.

3. The arc finder must employ rules to reject spurious detections such as diffraction spikes from bright stars or edge-on disk galaxies.

4. The arc finder must be able to process many images in a reasonable amount of time.

Here we describe an algorithm for identifying giant arcs - the arcs we are most interested in analyzing in this work. The algorithm was designed to reasonably comply with the above criteria. The parameters that define what we consider to be a giant arc, such as the minimum length and length-to-width ratio, are presented in Section 2.3.2. Figure 2.1 shows a flowchart of the steps involved in the algorithm and summarizes its key components. The detailed descriptions of the various steps that comprise the algorithm can be found in appendices A through F.

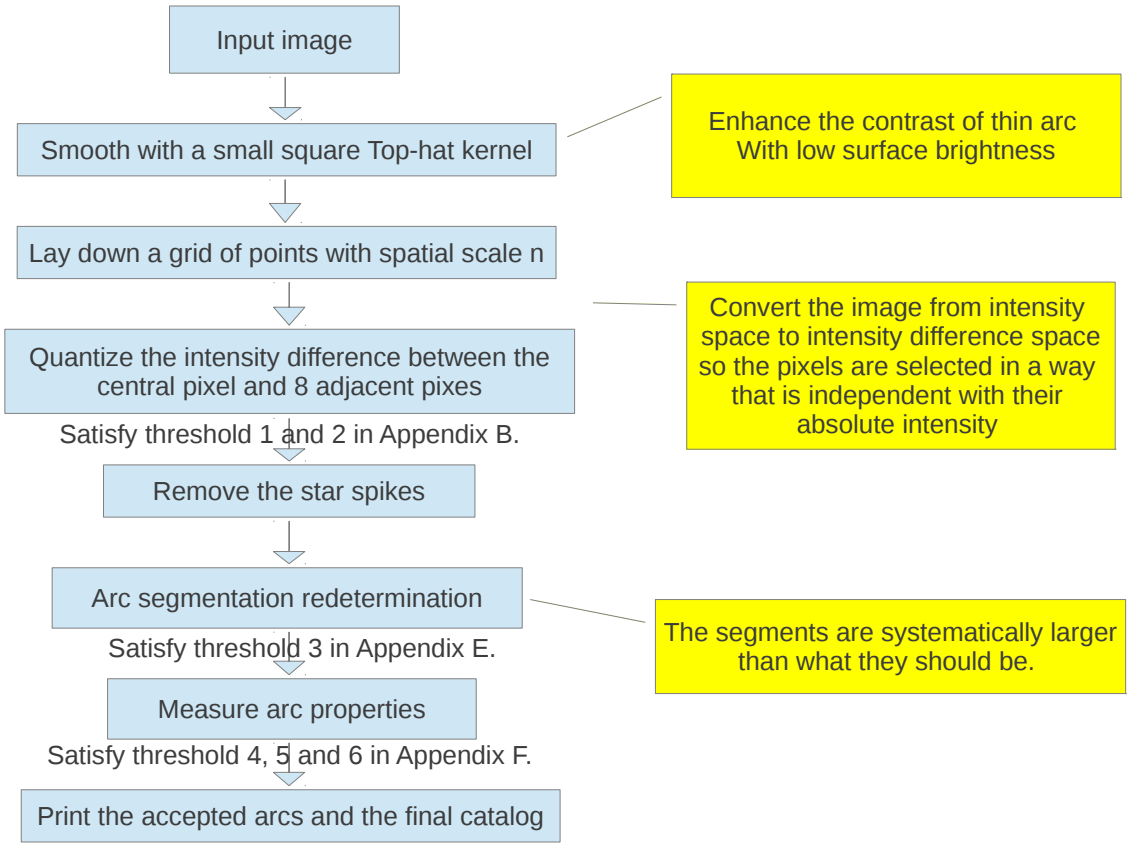


Figure 2.1: Flowchart of the arc-finding algorithm.

2.3 arc-finder Testing and Performance

2.3.1 Simulating Arcs

To compute the true arc abundance from the detected one, we must quantify the incompleteness and false positive rate of the arc-finder using a combination of actual and simulated datasets. The most robust way to simulate arcs is via ray-tracing in which light rays from objects in a source plane are shot towards the observer, deflected by the lens plane, and projected onto the image plane (the "sky" as seen by the observer). To quantify the incompleteness and false positive rate, one needs to trace large number of simulated arcs which is often very CPU intensive. Moreover, we need full control of all the input parameters of simulated arcs to perform the tests efficiently, and this becomes difficult to do solely by ray-tracing objects that are placed randomly on the source plane.

Furlanetto et al. (2013) use a different approach to simulate the arcs. Their basic idea is to represent an arc as a curved ellipse with its main axis being a segment of a circle. The model arc is then superposed directly on an image at various locations. The arc's shape is set by various parameters (e.g., length, width, curvature and orientation) chosen to mimic the shapes of real lensed galaxies. The surface brightness distribution is set using a Sersic law profile. The intensity parameters include the Sersic index and the intensity at the center of the arc, which allows one to assign any magnitude or the total flux to the

simulated arcs. However, this simple analytic prescription does not precisely reproduce the properties of real arcs. For example, the "painted-on" arcs tend to have a deficit of surface brightness at their long ends, which can result in shape measurement biases, especially for faint arcs. For a robust comparison between the real data and simulations, the "painted" arc method falls short of the fidelity that is required.

We adopt a hybrid approach to simulate the arcs: simulate a representative set of arcs via ray-tracing with a range of l/w ratios and surface brightnesses and then "paint" these template arcs onto the background images. This approach keeps the advantages of both methods: realistic arc rendering and fast performance. First, we perform ray-tracing by using simulated cluster lens with a NFW profile and a simulated background source with a Sersic profile. Second, we fine tune the distance from the source to the caustic line of the lens and carefully measure the l/w ratio of the formed arcs. We keep those arcs with l/w ratio that are closest to integer values as templates, as shown in Figure 2.2. We then create many additional simulated arcs by arbitrarily rotating the template images and by adjusting the total flux as desired. These arcs are then inserted into both simulated and actual CLASH images for our arc-finder performance testing. A detailed discussion on the general detectability of arcs as a function of source properties can be found in Meneghetti et al. (2008).

In order to accurately determine the arc detection completeness, we must

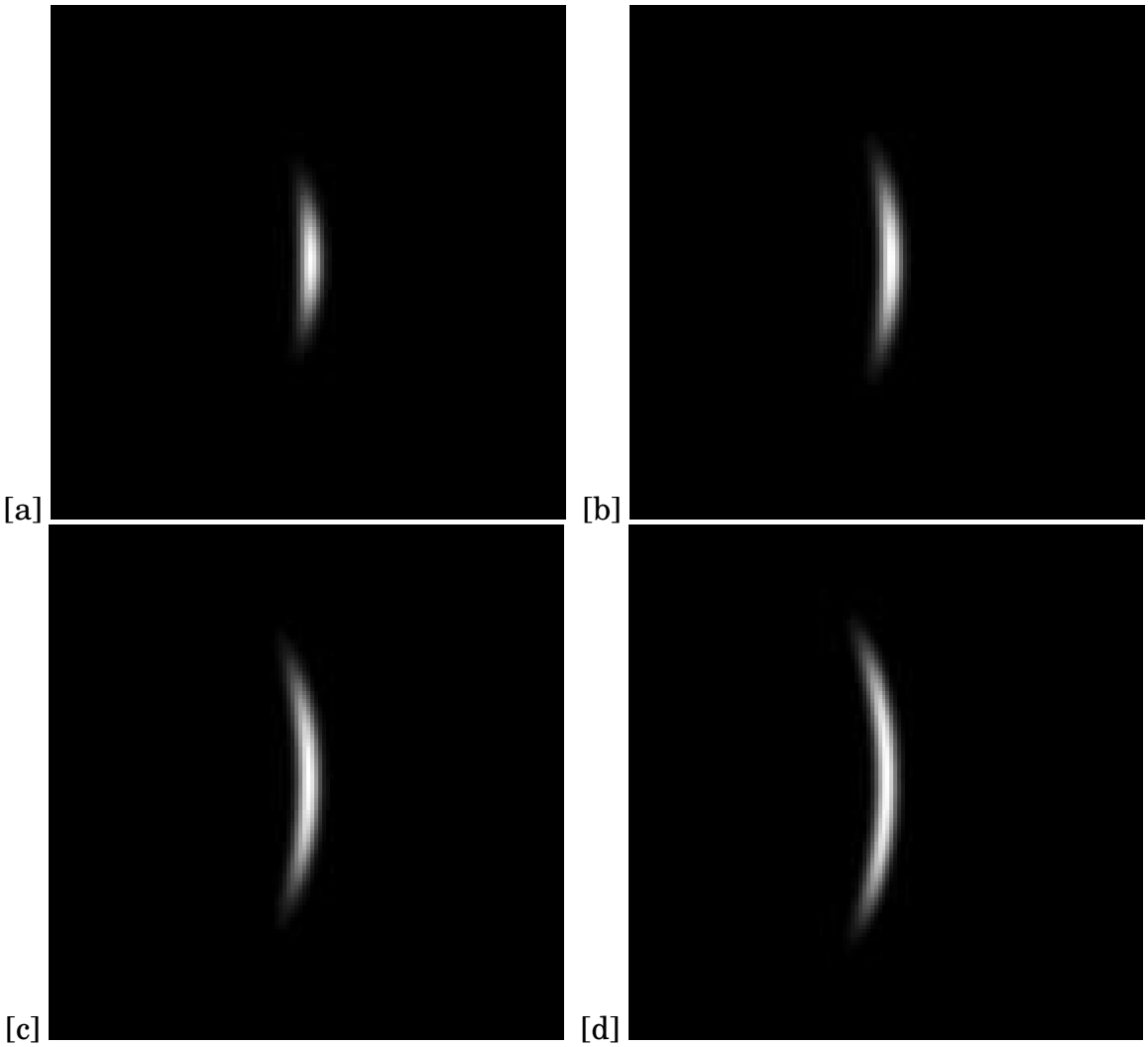


Figure 2.2: (a),(b),(c),(d) are arc templates with integer l/w ratios of 6, 7, 8, and 9, respectively, produced by the ray-tracing simulations.

account for the effects such as light contamination from cluster galaxies, variable sky background and instrumental noise. We accomplish this by inserting the arc templates into actual CLASH detection images (a weighted co-addition of all of the ACS and WFC3-IR images for a given cluster). An example of a CLASH detection image, with the brightest cluster galaxy subtracted out, is shown in Figure 2.3. We simulate a total of 14700 arcs spread over 7 different l/w values and 7 different total flux values. For the purpose of computing a measure of the algorithm's detection completeness, we only look at the fraction of simulated arcs that are detected, even though we are inserting the simulated arcs into real cluster data with real arcs. The completeness is then just the ratio of the number of the simulated arcs detected to the total number of arcs simulated. The inverse of this ratio, $f_{incom} = N_{sim}/N_{det,sim}$, is then the multiplicative factor that we will apply to any raw arc count to correct for incompleteness.

We also utilize the F814W CANDELS (Grogin et al., 2011; Koekemoer et al., 2011) images in the false positive rate test because there are no strong lensed sources in the CANDELS fields. We select images from the CANDELS "Wide programs" (e.g. the UDS and COSMOS fields), which have the similar total exposure times as the CLASH detection images ($\sim 50000s$) and split these mosaics into smaller images that match the angular size of the CLASH co-added images. We run the arc-finder on the CANDELS data and compute the

surface density of detections as functions of both the l/w threshold and the total arc length, l. This comprises the basis for our false positive correction function.

We use the CLASH data with simulated arcs to measure our arc-finder completeness as a function of arc length, l/w ratio, and arc signal-to-noise ratio (SNR). Here we define the SNR of arc as: $\sum I_i / \sqrt{\sum(\sigma_{i,bn}^2 + I_i)}$, where I_i is the intensity from the source at pixel i and $\sigma_{i,bn}^2$ is the combined variance due to the sky background and all sources of detector noise at pixel i. In our completeness test, the total flux of each drawn arc is adjusted to match the assigned SNR value. We use the CANDELS images to assess the arc-finder false positive rate. Figure 2.4 shows the completeness versus the l/w detection threshold, $(l/w)_{thr}$, at $S/N = 3, 10$. The completeness remains at a high level ($> 80\%$ for $S/N = 10$) when the $l/w \geq (l/w)_{thr}$; Figure 2.5 shows the false positive rate in the CANDELS data as a function the l/w detection threshold when the minimum arc length is set to $2''$. The detected number of false positives is slightly above 10 arcmin^{-2} at low $(l/w)_{thr}$, while it decrease rapidly as $(l/w)_{thr}$ increases. Figure 2.5 shows the false positive rate as a function of the length of the objects when the l/w threshold is set to 7. The number of the false positive detections peaks in the length bin $5'' \leq l < 6''$. The spurious detections can, thus, be suppressed if we adopt a minimum length threshold of $l \geq 6''$.

We have not applied this minimum length threshold to our completeness test because the identification of the arcs does not depend on the length (only

depends on l/w and S/N). Moreover, the intensity gradient along the ridge line of the arc should be smaller than that in the perpendicular direction, which mean that the length measurement should be more immune to the noise effects. To test that, we measure the ratio of the detected length to the true length of the simulated arcs. Figure 2.6 shows the distribution of the ratio at three different S/N levels. The dashed lines indicate the median value of the ratio. We can see that both the distribution and the median value remain statistically similar at different S/N levels.

2.3.2 Determination of the Optimal l/w Detection Threshold

In previous studies, the l/w detection threshold is typically set to 7.5, 8 or 10. Generally, the reason to set a high l/w threshold is to avoid the inclusion of highly elliptical and edge-on spiral galaxies into the arc sample. In general, the lower l/w threshold one uses, the more contamination one gets. Hence it is desirable to find a l/w threshold that maximizes the completeness level and minimizes the false positive rate. We now use our measured estimates of the completeness and false positive rate as a function of the minimal l/w to identify the optimal l/w threshold to use in the construction of our final arc catalog. We do this by identifying the smallest l/w threshold at which the surface density

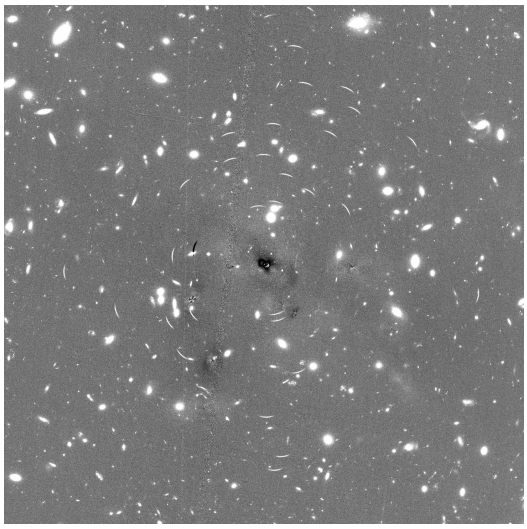
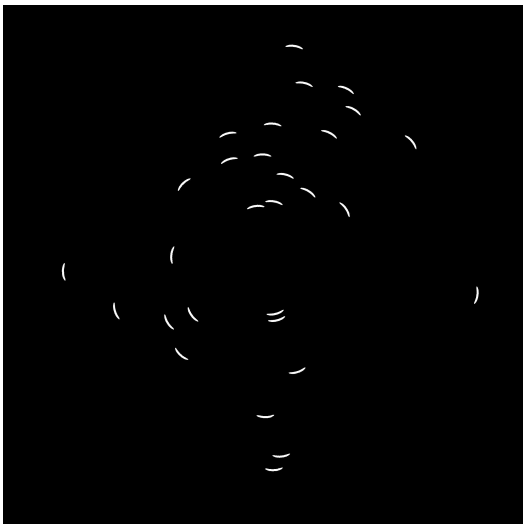
[a]  [b] 

Figure 2.3: (a) CLASH detection image for Abell 1423 shown with 30 "painted" arcs with $l/w = 7$. The brightest cluster galaxy and a handful of its satellites are first subtracted off before the arc-finder is run. (b) The arcs that are detected are shown. The FOV of both images is $2.7' \times 2.7'$.

of detected simulated arcs, N_{det}, exceeds the surface density of false positive detections, N_{fpr}, by a factor of 5 or more. The results of this test are shown in Figure 2.6. We find that the ratio $N_{det}(\geq (l/w)_{thr})/N_{fpr}(\geq (l/w)_{thr})$ is always larger than 5 when the l/w detection threshold is larger than 7. We thus adopt the l/w detection threshold of 7 in our analysis of the arc abundance. The false positive rate for $l/w \geq 7$ and an arc length threshold $l \geq 6''$ is 1.5 ± 0.4 arcmin^{-2}. We use this false positive rate to correct our corresponding raw arc counts.

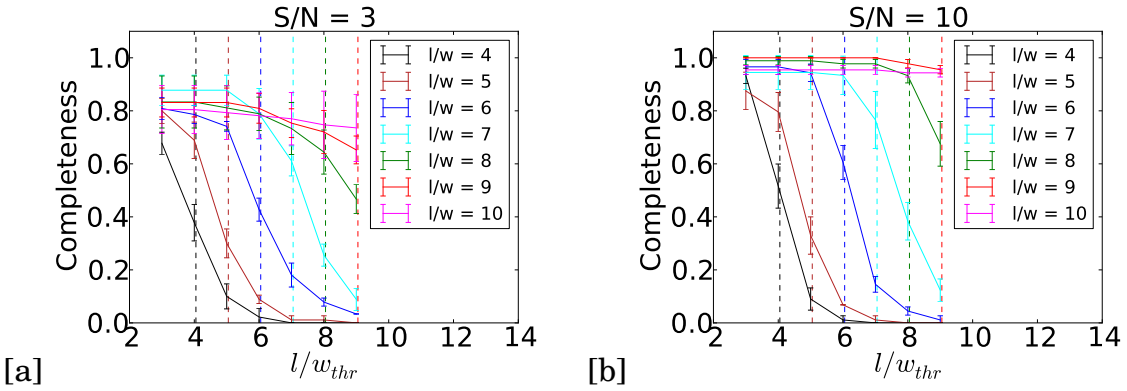


Figure 2.4: Figure (a), (b) show the completeness as a function of the l/w threshold for 7 different true l/w ratios at two different S/N levels. The dashed lines indicate the $l/w = (l/w)_{thr}$ and the error bars denote the 1σ rms error.

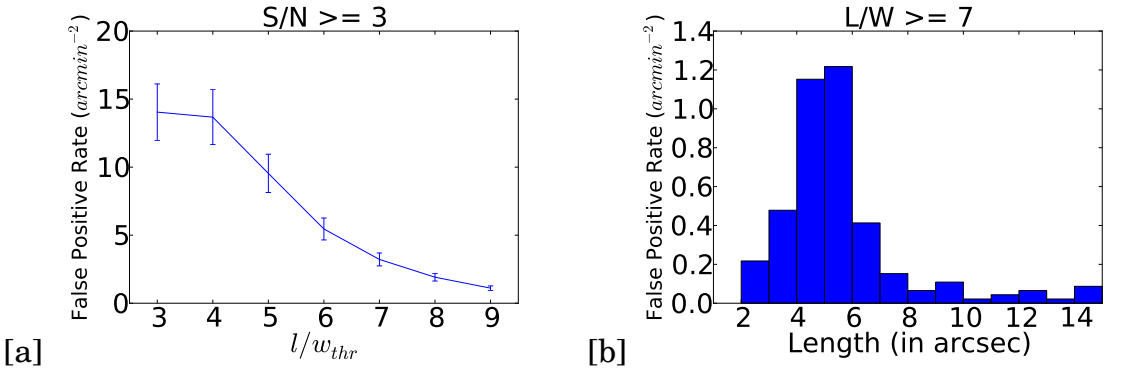


Figure 2.5: (a) The false positive rate per unit area as a function of the detected l/w threshold for arcs with SNR ≥ 3. (b) The false positive rate per unit area as a function of the arc length for arcs with $l/w \geq 7$. Results based on running the arc-finder on CANDELS data. The error bars denote the 1σ rms error.

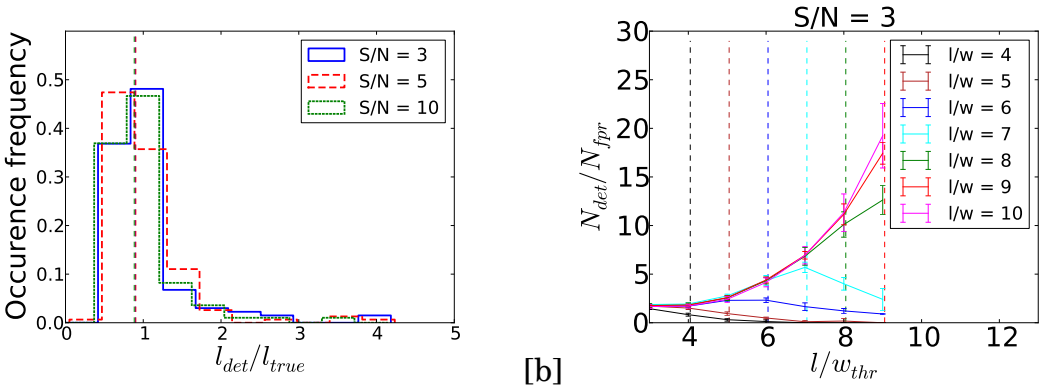


Figure 2.6: Figure(a) shows the distribution of the ratio of the detected length to the true length at three different S/N levels. The colored dashed lines denote the median values of the distribution; figure(b) shows the ratio of the number of detections per unit area to the number of false positive rate per unit area as a function of the detected l/w threshold. The dashed lines indicate the $l/w = (l/w)_{thr}$ and the error bars denote the 1σ rms error.

2.3.3 l/w Elongation Bias, Incompleteness and False Positive Rate Correction

There are three statistical corrections we need to apply to the raw counts of the giant arcs. First, the detected l/w is not equal to the true l/w. The background noise and/or the segmentation boundaries of a detected object may systematically affect the determination of the l/w ratio. We need to determine how the detected l/w ratio deviates from the true l/w ratio at different S/N levels, and correct for this elongation bias in a statistical sense. For example, as shown in Figure 2.7, the detected l/w ratio of arcs can be biased high by image noise, as the noise tends to make arcs appear thinner than they actually are. Second, we need to apply the incompleteness correction (presented above) as

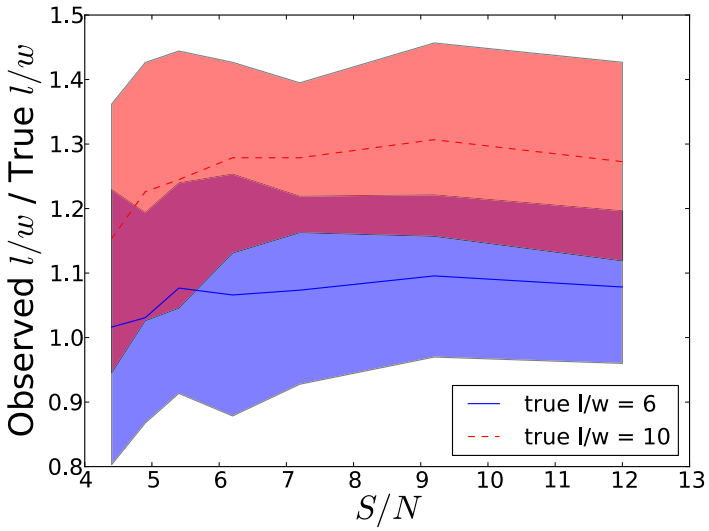


Figure 2.7: The ratio of the observed l/w to true l/w as a function of the arc signal-to-noise ratio, S/N. The height of the shaded regions denotes the 1σ rms errors on the ratio.

there will always be some real arcs that are missed by our detection algorithm. Third, we need to apply a false positive correction as there are always some objects misidentified by the arc-finding algorithm. We apply all these three corrections in deriving the final observed and simulated arc abundances.

To compute the l/w elongation bias correction, we collect all the detected arcs with measured $l/w \geq 6.5$ [1], and assign them to one of the three bins: 6.5 $\leq l/w <$ 7.5, 7.5 $\leq l/w <$ 8.5 and $l/w \geq$ 8.5. We also assign their corresponding true l/w ratios into one of the three bins: $l/w = 7$, $l/w = 8$ and $l/w \geq 9$. We further split each bin into three sub-bins by their S/N ratios: $S/N < 5$, $5 \leq$

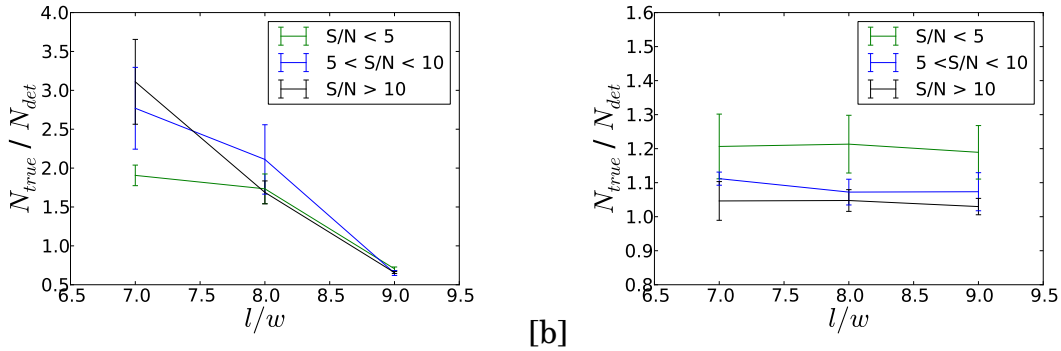
[1] In practice, we set the l/w threshold to be 6.5 instead of 7. The bias correction is then done by comparing the arcs with detected ratios in the range $6.5 \leq l/w < 7.5$ to the number with true $l/w = 7$.

$S/N < 10$ and $S/N \geq 10$. We then calculate the mean value and standard deviation of the correction factor for the elongation bias $f_{bias} = N_{true}/N_{det}$, where N_{true} and N_{det} are the number of simulated arcs and detected arcs in each bin, respectively.

The true arc count, N_{tru}, is then computed as follows:

$$N_{true} = \sum_i N_{det,i} \times f_{bias,i} \times f_{incom,i} - N_{false}$$
$$\sigma_{true} = \sqrt{N_{true} \times \sum_i [(\frac{\sigma_{bias,i}}{f_{bias,i}})^2 + (\frac{\sigma_{incom,i}}{f_{incom,i}})^2] + \sigma_{false}^2} \quad (2.1)$$

where $N_{det,i}$ is the observed number of arcs in each bin and i goes over all the bins. As shown in Figure 2.8, most of the measured l/w are biased high, especially for the arcs with intrinsically low l/w. The completeness remains above 80% for all the cases. Here, biased high means that arcs with "true" low l/w have their l/w values systematically overestimated. The mean ratio of the observed l/w to the "true" l/w also appears to be dependent on the true l/w ratio as shown in Figure 2.7.



[a] [b]

Figure 2.8: All the detected arcs with $l/w \geq 6.5$ are assigned into three l/w bins (horizontal axis). (a) shows the elongation bias correction factors in three l/w bins; (b) shows the incompleteness correction factors in three l/w bins. The error bars denotes the 1σ rms error.

2.3.4 Comparison of arc-finder's Performance with a Previous Code

We compare the arc detection efficiency of our arc-finder to that of the only publicly available arc-finding code from Horesh et al. (2005). We simulate a large amount of arc with different l/w ratios and draw gaussian random noise onto the arcs to produce simulated arc images with 7 different S/N levels. We run both arc-finding algorithms on these simulated data sets.

Figure 2.9 shows the detection rate versus the arc S/N ratio level for arcs with true $l/w = 7$, 10, at a detection threshold $l/w \geq 7$. We have computed the S/N ratio for detected arcs found using each of the algorithms using the definition given in §2.3.1. For the bright arcs ($S/N > 10$), the detection rates for both arc-finders remains high ($> 90\%$); for faint arcs ($5 \leq S/N < 10$), the Horesh

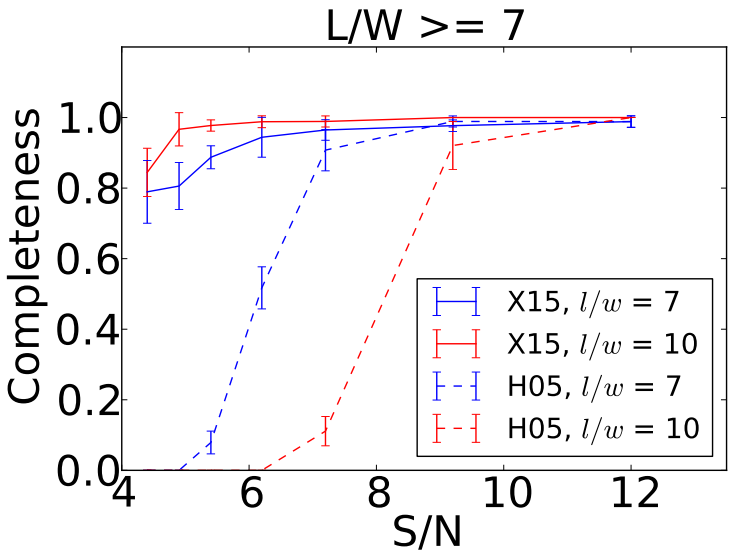


Figure 2.9: Comparison of the arc detection completeness between our arc-finder (solid curves) and the Horesh et al. (2005) arc-finder (dashed curves) for arcs with $l/w = 7$, 10. The error bars denote the scatter. For bright arcs with $S/N > 10$ both arc-finders maintain a high detection rate ($> 90\%$); while for faint arcs with lower S/N levels, our arc-finder exhibits considerably higher detection efficiency. The error bar denotes the 1σ rms error.

et al. (2005) arc-finder's detection rate drops rapidly, while our algorithm's detection efficiency remains higher than 90%; for very faint arcs ($S/N < 5$), our detection rate drops to about 80%. The advantage of our intensity-gradient based arc-finding algorithm is nicely demonstrated in Figure 2.9, especially for the detection of large arcs with low-surface brightness.

2.4 Analyses of the CLASH Data

2.4.1 Arc Statistics for the CLASH Sample

The CLASH observations for each cluster consist of 16 broadband images (spanning the range 0.23μm - 1.6μm) using the $WFC3/UVIS$, $WFC3/IR$, and ACS/WFC instruments onboard HST. The cluster properties are listed in Table 2.1. We run our arc-finder on the detection (ACS+$WFC3/IR$) image created for each cluster. We detect a raw total of 187 arcs with $l/w \geq 6.5$ in 20 X-ray selected CLASH clusters. After applying our minimum arc length criterion $l \geq 6''$, the arc count drops to 81 giant arcs selected from the 20 X-ray selected CLASH clusters. Correcting for the elongation bias and incompleteness brings the total number of detected arcs in 20 X-ray selected clusters is 104 ± 12. After further correcting for the false positive rate, we find a lensing efficiency of 4 ± 1 arcs per X-ray selected cluster. Throughout this paper, the lensing efficiency denotes the number of arcs per cluster. There are 28 arcs with $l/w \geq 6.5$ and $l \geq 6''$ detected in the five high-magnification CLASH clusters, corresponding to a mean value 5 ± 1 arcs per cluster after all corrections are applied. Figure 2.10 shows the distributions of number of arcs per cluster for the X-ray selected cluster sample and the high magnification cluster sample. Figure 2.11 shows the comparison of the detection images with the raw output of the arc-finder with $l/w > 7$ for five CLASH clusters.

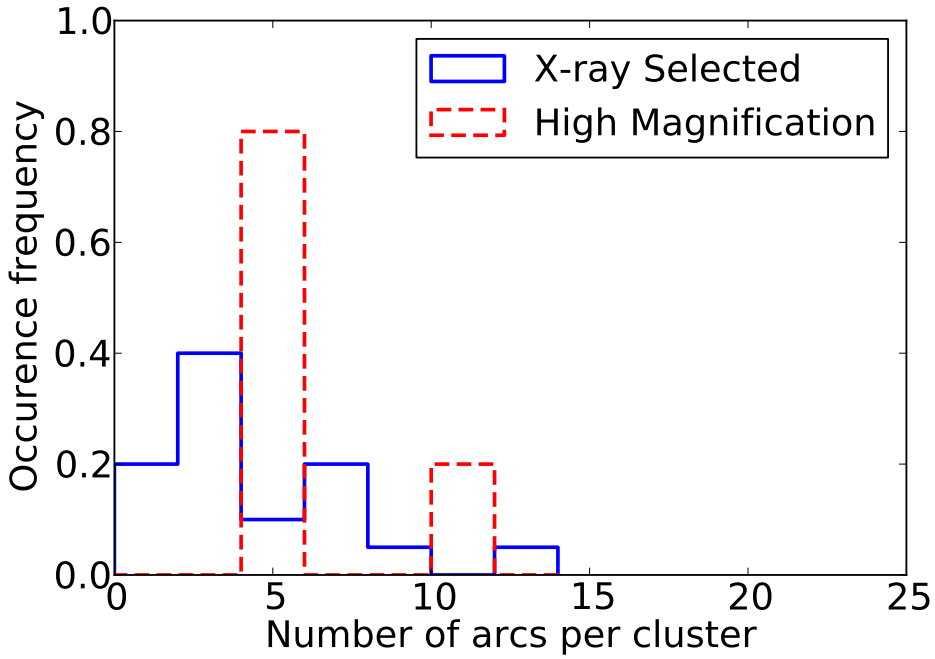


Figure 2.10: The occurrence frequency of arcs per cluster for 20 X-ray selected CLASH clusters in blue and for 5 high lens magnification subsample of CLASH clusters in red.

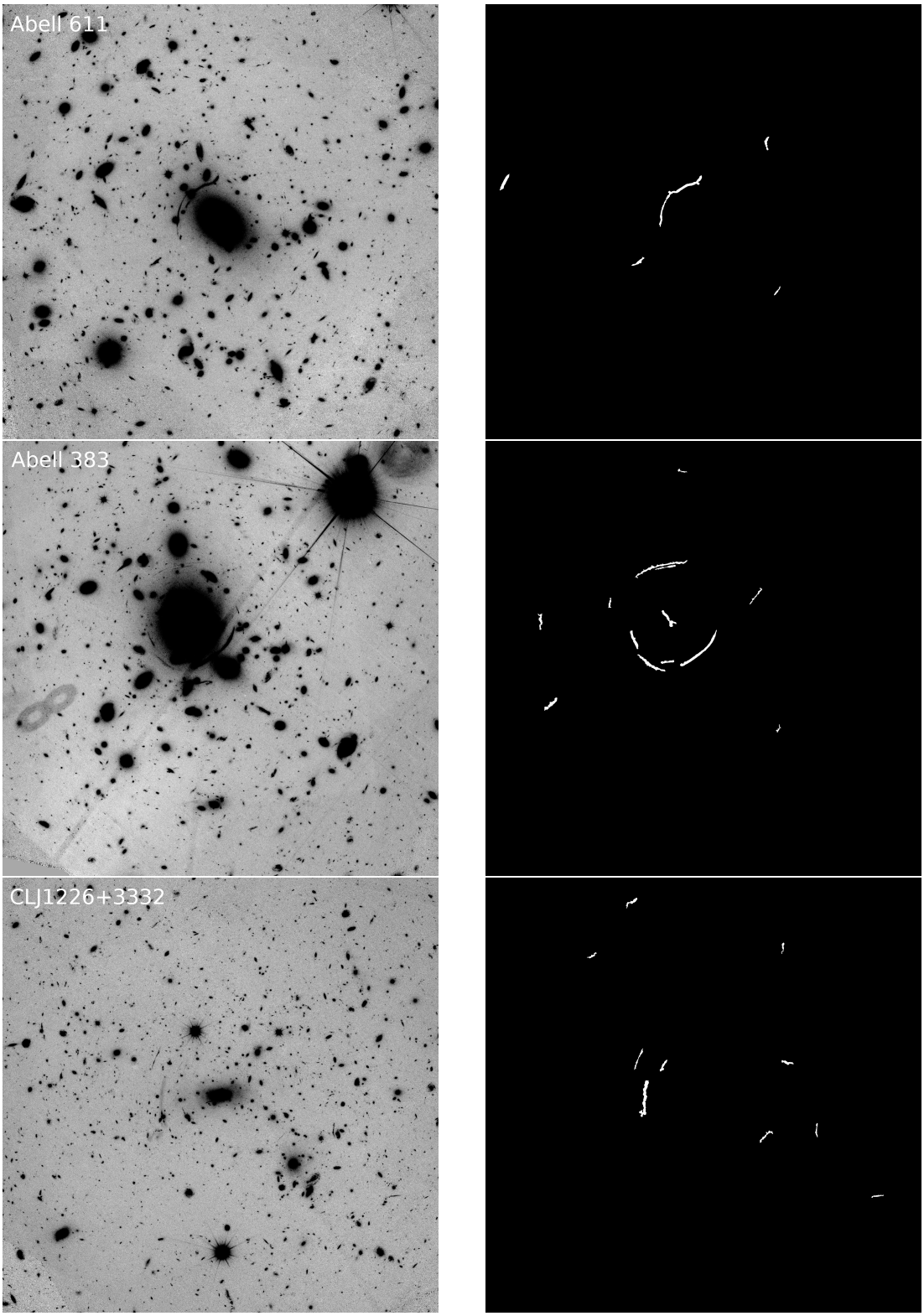


Figure 2.11: The detection images and segmentation maps of five CLASH clusters.

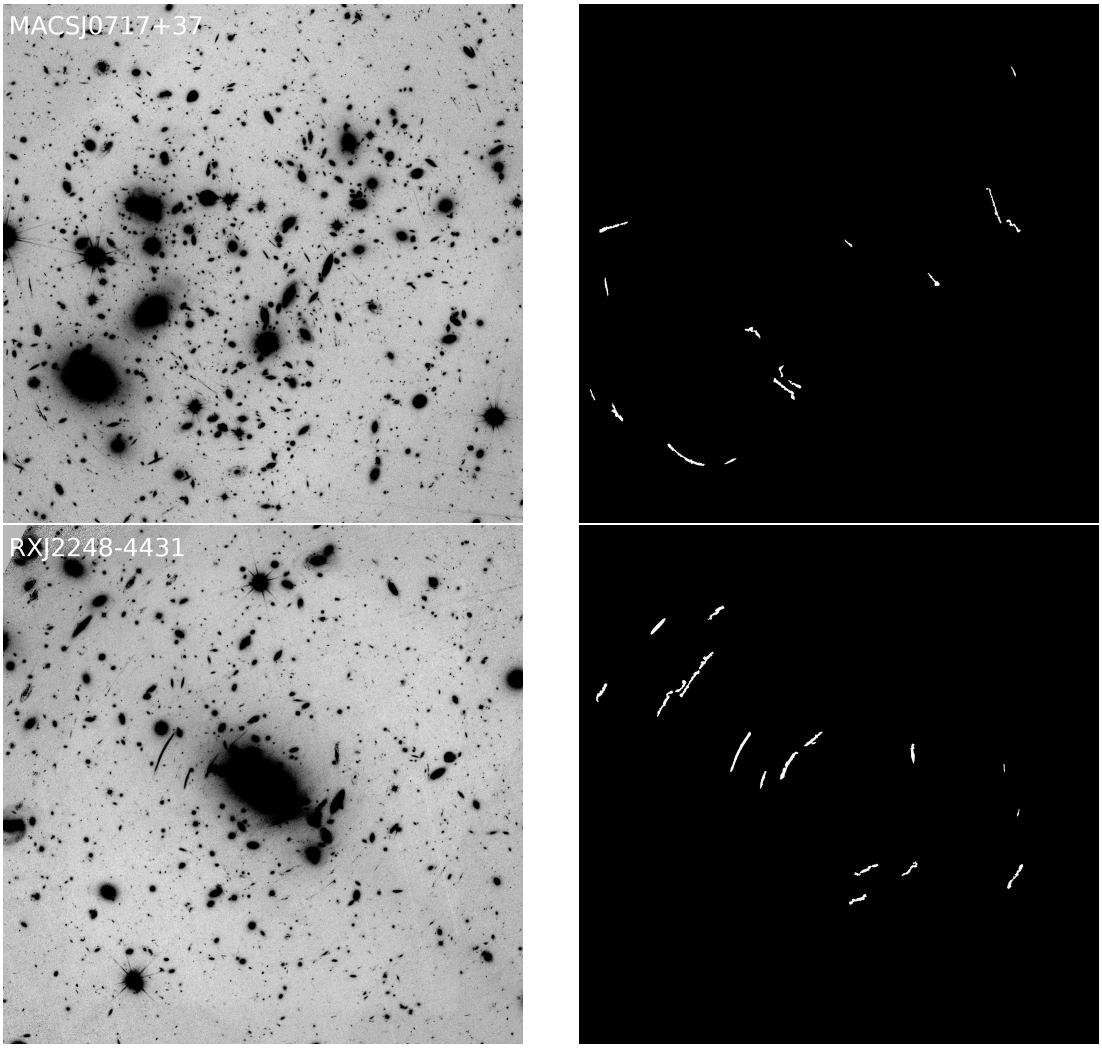


Figure 2.11: Left panel shows the detection images of five CLASH clusters; right panel shows the raw output maps produced by the arc-finder with $l/w > 7$.

Table 2.1: The CLASH cluster sample.

Cluster	$\alpha_{\rm J2000}$	$\delta_{\rm J2000}$	$z_{\rm Clus}$	M_{200c} $[10^{15}M_{\odot}/h]$
X-ray Selected Clusters:				
Abell 209	01:31:52.57	−13:36:38.8	0.206	0.95 ± 0.07
Abell 383	02:48:03.36	−03:31:44.7	0.187	0.87 ± 0.07
MACS0329.7-0211	03:29:41.68	−02:11:47.7	0.450	0.73 ± 0.10
MACS0429.6-0253	04:29:36.10	−02:53:08.0	0.399	0.80 ± 0.14
MACS0744.9+3927	07:44:52.80	+39:27:24.4	0.686	0.70 ± 0.04
Abell 611	08:00:56.83	+36:03:24.1	0.288	0.85 ± 0.05
MACS1115.9+0129	11:15:52.05	+01:29:56.6	0.352	0.90 ± 0.09
Abell 1423	11:57:17.26	+33:36:37.4	0.213	...
MACS1206.2-0847	12:06:12.28	−08:48:02.4	0.440	0.86 ± 0.11
CLJ1226.9+3332	12:26:58.37	+33:32:47.4	0.890	1.56 ± 0.10
MACS1311.0-0310	13:11:01.67	−03:10:39.5	0.494	0.46 ± 0.03
RXJ1347.5-1145	13:47:30.59	−11:45:10.1	0.451	1.16 ± 0.19
MACS1423.8+2404	14:23:47.76	+24:04:40.5	0.545	0.57 ± 0.10
RXJ1532.9+3021	15:32:53.78	+30:20:58.7	0.345	0.53 ± 0.08
MACS1720.3+3536	17:20:16.95	+35:36:23.6	0.391	0.75 ± 0.08
Abell 2261	17:22:27.25	+32:07:58.6	0.224	1.42 ± 0.17
MACS1931.8-2635	19:31:49.66	−26:34:34.0	0.352	0.69 ± 0.05
RXJ2129.7+0005	21:29:39.94	+00:05:18.8	0.234	0.61 ± 0.06
MS2137-2353	21:40:15.18	−23:39:40.7	0.313	1.04 ± 0.06
RXJ2248.7-4431	22:48:44.29	−44:31:48.4	0.348	1.16 ± 0.12
High Magnification Clusters:				
MACS0416.1-2403	04:16:09.39	−24:04:03.9	0.420	...
MACS0647.8+7015	06:47:50.03	+70:14:49.7	0.584	...
MACS0717.5+3745	07:17:31.65	+37:45:18.5	0.548	...
MACS1149.6+2223	11:49:35.86	+22:23:55.0	0.544	...
MACS2129.4-0741	21:29:26.06[a]	−07:41:28.8[a]	0.570	...

2.4.2 The Arc Redshift Distribution

We determine the photometric redshift distribution of the lensed background galaxies detected by our algorithm using the photometric redshifts derived with the Bayesian-based BPZ package (BPZ; Benítez (2000, 2004); Coe et al. (2006)). Spectral energy distribution (SED) templates are redshifted and fit to the observed photometry. The BPZ code adopts a prior that the empirical likelihood of redshift is a function of both galaxy magnitude and galaxy morphological type (e.g., bright and/or elliptical galaxies are rare at high redshift). We used 11 SED templates originally from PEGASE (Fioc & Rocca-Volmerange, 1997) that have been recalibrated based on photometry and spectroscopic redshifts of galaxies in the FIREWORKS catalog (Wuyts et al., 2008). We obtain the photometric redshift distribution of all the detected arcs and find that they have a median photometric redshift $z_s = 1.9$. We also find that there is a significant fraction of arcs with $z_s \sim 3$ (34% of the detected arcs have photometric redshift larger than 3). Figure 2.12 shows the arc number counts as a function of redshift before and after correcting for the measurement bias, incompleteness and false positive rate. To compute the photometric redshift distribution of our arc sample, we sum up the individual posterior redshift probability distributions of each detected arc. The mean uncertainty of the photometric redshifts in CLASH is $\sigma_z \sim 0.03(1+z)$ and, thus, we sample the probability distribution using the bin size $\Delta z = 0.4$, which is twice as large as the uncertainty

of the arc with highest redshift. The summed distribution prior to correction for our selection function and elongation bias is shown as the blue line in Figure 2.12. To correct for the selection bias, incompleteness and false positive rate, we re-sum the probability distribution for each arc after first multiplying by the appropriate correction factors. The fully corrected redshift distribution, derived in this way, is shown by the red line in in Figure 2.12. Figure 2.13 also lists the distribution of arc S/N ratio, arc AB magnitude in F814W filter, arc l/w ratio, and the normalized angular distance of the arc from the cluster center. Table 2.2 lists the properties of all the detected arcs in 20 X-ray selected sample, including the equatorial and pixel coordinates, length, l/w ratio, radial distance from the arc center to the cluster center, the normalized radial distance by r_{200}, the photometric redshift and the AB magnitude in the F814W band. In Table 2.2, we do not exclude the objects with photometric redshifts that are significantly smaller than the corresponding cluster redshift. Such probable foreground sources are considered to be false positive detections. We eliminate false positive detections statistically when we calculate the arc redshift distribution in the CLASH sample.

Table 2.2: Detected arcs and their physical properties. RD = radial distance from the arc center to the cluster center in the unit of arcsecond; IDs with † denote the false positive detection.

Cluster	Arc ID	RA	DEC	x	y	l ($''$)	l/w	RD ($''$)	RD/r_{200}	z	AB mag (F814W)
Abell1423	1	179.33	33.60	2204.00	1905.00	6.82	9.67	43.09	1.09	0.62	22.61
	†2	179.30	33.62	3553.00	3227.00	6.14	7.61	82.69	2.10	0.00	24.29
Abell209	1	22.96	-13.61	2760.00	2738.00	11.90	10.00	23.16	0.62	3.50	21.73
Abell2261	†1	260.59	32.12	3542.00	1698.00	7.08	8.89	84.98	1.84	0.00	24.39
	†2	260.61	32.12	2505.00	1994.00	8.46	10.55	32.18	0.70	0.33	23.01
	3	260.62	32.13	2165.00	2317.00	6.72	7.92	24.54	0.53	3.54	23.75

To be continued

Cluster	Arc ID	RA	DEC	x	y	l ($''$)	l/w	RD ($''$)	RD/r_{200}	z	AB mag (F814W)
	4	260.60	32.13	2902.00	2395.00	10.55	14.29	26.77	0.58	1.80	22.07
	†5	260.64	32.13	1324.00	2620.00	13.06	23.34	76.98	1.67	0.27	23.79
	6	260.61	32.15	2577.00	3516.00	6.64	10.20	66.94	1.45	1.35	23.22
Abell383	1	42.03	-3.54	1620.00	2246.00	6.80	7.03	54.43	1.30	0.73	25.74
	2	42.02	-3.53	2181.00	2462.00	12.36	14.63	17.70	0.42	4.22	22.94
	3	42.01	-3.53	2553.00	2616.00	19.23	22.41	15.27	0.36	0.89	20.15
	4	42.02	-3.53	2114.00	2579.00	7.32	7.91	15.79	0.38	3.12	23.20
	5	42.03	-3.53	1566.00	2735.00	6.33	8.56	49.75	1.18	2.46	25.06
	†6	42.01	-3.53	2314.00	2691.00	8.43	9.01	1.04	0.02	0.30	21.93
	7	42.00	-3.53	2828.00	2893.00	7.48	17.11	34.96	0.83	6.31	23.92
	8	42.01	-3.52	2281.00	3035.00	20.42	31.37	22.65	0.54	5.00	23.33
	9	42.01	-3.52	2281.00	3024.00	7.87	16.59	21.94	0.52	3.24	24.81
Abell611	1	120.24	36.06	2335.00	2665.00	25.66	30.64	15.35	0.40	1.12	20.44
	†2	120.26	36.06	1358.00	2727.00	7.02	7.37	75.89	2.00	0.27	19.75
CLJ1226	1	186.74	33.54	2838.00	2253.00	6.30	9.52	26.72	1.52	2.79	25.90
	2	186.75	33.54	2164.00	2394.00	14.22	12.37	23.10	1.31	2.30	24.96
	3	186.75	33.55	2144.00	2742.00	8.47	19.19	28.47	1.62	3.45	23.79
MACS0329	1	52.41	-2.19	3049.00	2869.00	7.96	9.55	43.27	1.56	1.03	23.28
	2	52.42	-2.18	2685.00	3153.00	6.20	9.45	44.66	1.61	3.35	24.25
MACS0429	1	67.40	-2.90	2752.00	1709.00	6.48	8.01	52.93	2.08	1.67	24.22
	2	67.40	-2.89	2499.00	2200.00	10.69	12.24	18.83	0.74	1.35	21.79
MACS0744	†1	116.22	39.44	2682.00	1488.00	7.00	7.56	66.12	3.43	0.41	20.22
	2	116.23	39.45	2128.00	2078.00	6.41	7.74	36.39	1.89	4.79	23.85
	†3	116.20	39.45	3147.00	2191.00	7.32	7.32	45.92	2.38	0.14	18.48
	4	116.23	39.46	1969.00	2581.00	6.75	9.40	35.47	1.84	4.73	23.99
	5	116.23	39.46	2033.00	2625.00	6.10	8.37	32.04	1.66	4.41	24.43
	6	116.20	39.46	3477.00	2652.00	7.56	14.86	63.92	3.32	4.11	23.76
	7	116.21	39.46	2839.00	2632.00	6.77	7.71	23.47	1.22	1.17	20.34
MACS1115	1	168.96	1.49	2602.00	2228.00	14.44	18.05	18.34	0.58	2.46	23.08
	2	168.98	1.50	1586.00	2402.00	9.53	15.25	60.03	1.90	1.76	24.83
	†3	168.97	1.50	2500.00	2374.00	7.76	7.60	7.74	0.25	0.42	20.91
	4	168.97	1.50	2355.00	2371.00	10.65	12.58	12.57	0.40	4.21	21.72
	5	168.96	1.51	2896.00	2911.00	14.52	16.16	37.20	1.18	4.12	22.91
	6	168.96	1.51	2578.00	3022.00	7.14	7.11	34.71	1.10	3.25	24.41
MACS1206	†1	181.55	-8.81	2247.00	2078.00	6.69	7.04	28.16	0.99	0.55	20.26
	2	181.54	-8.80	2790.00	2454.00	14.27	15.73	19.47	0.68	1.05	19.76
	3	181.54	-8.80	3292.00	2420.00	6.54	7.51	52.08	1.83	2.41	24.64
	†4	181.55	-8.80	2477.00	2438.00	11.97	10.88	0.92	0.03	0.49	23.77
	5	181.57	-8.80	1618.00	2471.00	8.39	12.32	56.79	1.99	1.56	23.19
	6	181.53	-8.79	3484.00	3134.00	8.36	8.57	78.90	2.77	0.72	19.30
MACS1311	1	197.75	-3.17	2881.00	2742.00	6.25	7.66	29.33	1.30	2.83	24.77
MACS1423	†1	215.93	24.07	3433.00	2050.00	9.84	17.73	66.78	3.12	0.00	23.74
	2	215.95	24.07	2606.00	2128.00	6.13	7.80	24.49	1.14	1.47	23.92
	3	215.94	24.07	2804.00	2120.00	8.80	15.30	30.97	1.45	2.57	23.94
	4	215.95	24.08	2378.00	2773.00	8.18	11.48	20.10	0.94	1.79	22.72
	5	215.95	24.09	2576.00	3211.00	7.70	11.46	47.03	2.20	3.16	24.47
MACS1720	1	260.06	35.60	2757.00	2042.00	8.33	11.23	33.32	1.20	4.38	24.29
	2	260.07	35.60	2621.00	2346.00	8.60	10.80	11.90	0.43	0.82	23.23
MACS1931	1	292.96	-26.59	2559.00	1917.00	9.71	11.49	37.90	1.38	3.55	24.23
MS2137	1	325.06	-23.66	2422.00	2713.00	14.81	12.99	14.59	0.46	1.77	21.78
	2	325.07	-23.65	2181.00	2903.00	10.75	14.08	33.37	1.05	1.71	23.95
	3	325.07	-23.65	2215.00	3046.00	10.93	12.95	39.93	1.26	1.97	23.86
RXJ1347	1	206.87	-11.77	3032.00	1776.00	12.76	21.51	55.55	1.82	1.64	23.58
	2	206.87	-11.76	3141.00	2064.00	9.07	11.79	46.58	1.52	2.43	21.57
	3	206.87	-11.75	3117.00	2872.00	7.02	10.12	43.42	1.42	4.28	24.63
	4	206.88	-11.75	2521.00	2951.00	6.82	7.51	29.92	0.98	0.78	21.46
	5	206.88	-11.74	2549.00	3162.00	7.83	13.50	43.50	1.42	3.78	24.22
RXJ1532	†1	233.22	30.34	2841.00	2088.00	7.75	9.30	34.16	1.25	0.27	22.25
RXJ2129	1	322.41	0.08	2890.00	2044.00	7.42	7.79	38.34	1.36	3.17	23.61
	†2	322.44	0.09	1394.00	2310.00	6.26	9.57	73.09	2.60	0.00	24.96
	3	322.42	0.09	2295.00	2528.00	7.20	7.83	13.81	0.49	1.55	22.65
RXJ2248	1	342.18	-44.54	2613.00	1950.00	6.15	6.89	36.42	1.31	3.08	24.18

To be continued

Cluster	Arc ID	RA	DEC	x	y	l ($''$)	l/w	RD ($''$)	RD/r_{200}	z	AB mag (F814W)
	2	342.16	-44.54	3371.00	2095.00	8.91	11.34	62.35	2.25	3.09	24.02
	3	342.18	-44.54	2622.00	2084.00	8.06	10.83	28.10	1.01	1.64	23.35
	4	342.17	-44.54	2830.00	2074.00	6.61	9.53	34.94	1.26	2.75	24.42
	5	342.19	-44.53	2227.00	2545.00	10.12	10.26	18.06	0.65	1.41	22.36
	6	342.19	-44.53	2039.00	2697.00	13.95	17.47	32.67	1.18	1.38	21.32
	7	342.17	-44.53	2853.00	2642.00	6.19	7.28	24.70	0.89	1.41	22.35
	8	342.19	-44.53	2343.00	2684.00	7.32	8.80	15.81	0.57	3.76	24.48
	9	342.20	-44.52	1664.00	2902.00	9.44	13.06	60.38	2.18	1.36	25.23
	†10	342.21	-44.52	1358.00	2943.00	6.42	7.33	79.70	2.87	0.00	21.79
	11	342.20	-44.52	1743.00	2930.00	10.16	14.05	56.68	2.04	2.79	24.33
	12	342.20	-44.52	1856.00	3108.00	6.95	7.97	57.66	2.08	1.96	24.89
	†13	342.21	-44.52	1629.00	3261.00	6.73	6.85	75.27	2.71	0.40	19.07
	14	342.20	-44.52	1909.00	3340.00	6.49	7.68	66.85	2.41	0.85	24.11

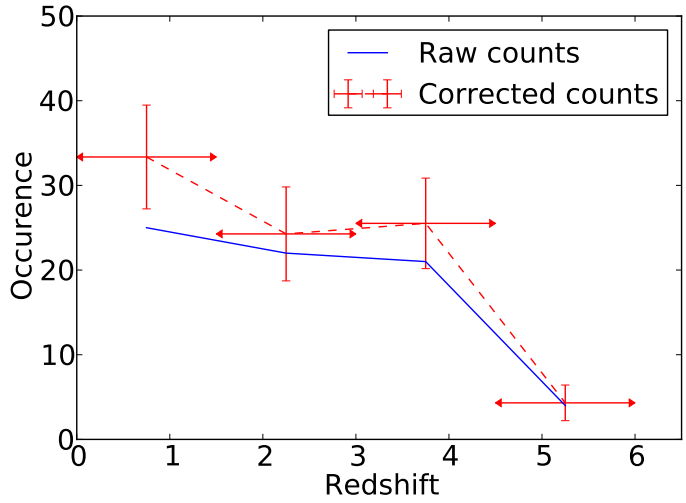


Figure 2.12: The photometric redshift distribution of the detected arcs in the CLASH X-ray selected sample. The blue solid line denotes the redshift distribution of the raw data counts, which is computed based on the full posterior probability distribution of the detected arcs; the red dashed line denotes the redshift distribution after the elongation bias, incompleteness and false positive correction, which is computed based on the corrected full posterior probability distribution. The error bar represents the 1σ Poisson error.

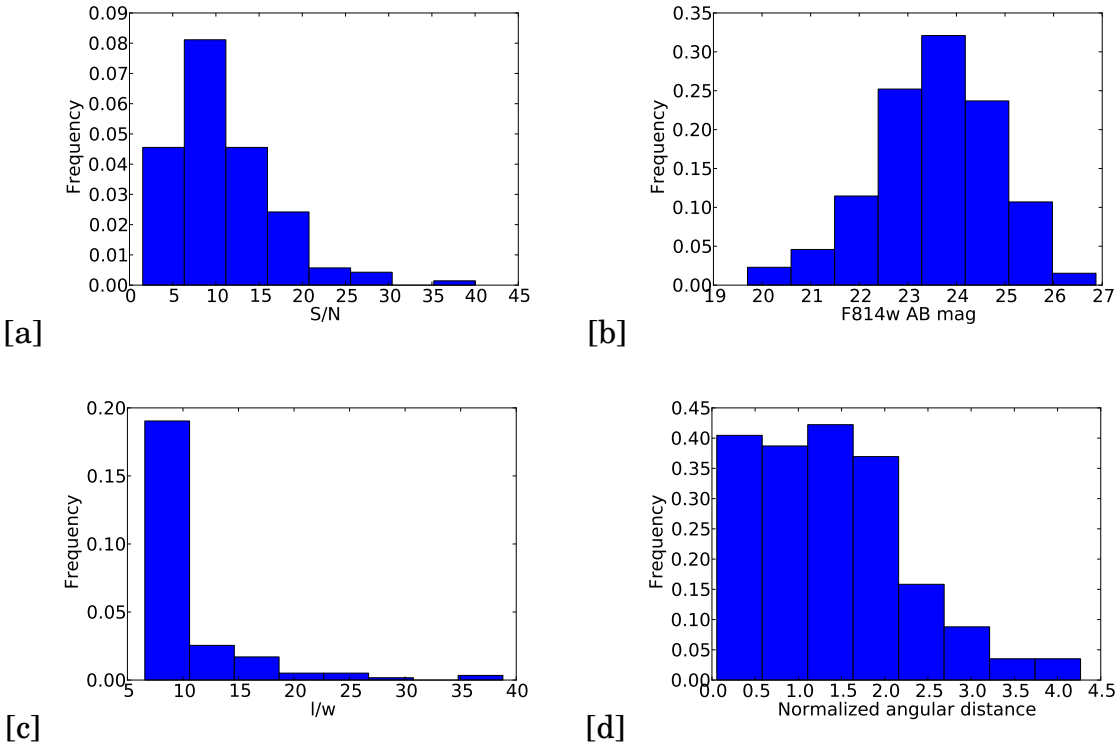


Figure 2.13: Panel (a),(b),(c),(d) show the distribution of S/N ratio, AB magnitude, l/w ratio and normalized angular distance RD/r_{200} of all the detected arcs in the CLASH sample.

2.5 MOKA Lensing Simulations

2.5.1 The MOKA Simulated Cluster Sample

In order to confirm or resolve the arc statistics problem, we require realistic model predictions to compare with the observed CLASH arc counts. In previous studies, mock clusters were selected from N-body simulations using either dark matter only (Wambsganss et al., 2004; Hilbert et al., 2007; Horesh et al., 2010) or dark matter with other ingredients (Puchwein et al., 2005; Gottlöber & Yepes, 2007; Rozo et al., 2008; Hilbert et al., 2008; Meneghetti et al., 2010). The simulated clusters were then projected onto the plane of the sky as viewed from various directions to create the 2-D mass models. However, the total number of clusters and/or their mass and redshift ranges used in these prior studies are not optimally matched to the CLASH dataset. Therefore, we generate a simulated cluster sample by running the publicly available MOKA package (Giocoli et al., 2012). MOKA uses simulation-calibrated analytical relations to describe the dark matter and baryonic content of clusters, which allows one to incorporate all the cluster properties that are relevant for strong cluster lensing. For example, for each halo, a triaxial NFW profile and a random orientation are assigned. The axial ratios are generated from the prescriptions of Jing & Suto (2002). The halo concentration, c, and its dependence on cluster mass, M, and redshift are modeled based on the $c - M$ relation of Bhattacharya et

al. (2013) The joint weak lensing + strong lensing analysis by Merten et al. (2015); Umetsu et al. (2014) indicates that the observed $c - M$ relation derived from the 20 X-ray selected CLASH clusters agrees with the relation presented in Bhattacharya et al. (2013). The scatter in the concentration value at a fixed mass is well-described by a Gaussian distribution instead of a log-normal distribution, with rms ~ 0.33. We adopt this scatter in our MOKA simulations. The dark matter substructures, the central brightest cluster galaxy (BCG) and adiabatic contraction are also incorporated into the MOKA generated models. MOKA is computationally efficient and is able to create a single simulated cluster lens model in a few CPU seconds on a personal computer by using a fast semi-analytic approach. The details of the code and its implementation can be found in Giocoli et al. (2012).

For our study, we create 640 mock clusters with the same mass and redshift range as the 20 X-ray selected CLASH clusters (32 different realizations for each of the corresponding mass and redshift values). In particular, the density profile of the main halo follows a NFW profile while the density profile of the subhalos is chosen to be truncated Singular Isothermal Sphere (SIS) profile; the spatial density distribution of the subhalos follows the measurement from numerical simulations by Gao et al. (2004); the mass resolution of the subhalos is $10^{10} M_{\odot}$. We calculate the deflection angle, convergence and shear fields for each projected mass distribution. The angular resolution of the simulated

cluster images is $0.065''$, which matches the pixel scale in the CLASH images.

2.5.2 Background Source Images and Ray-Tracing Method

To create the sky scene from the MOKA mass models, we follow a methodology similar to that in Horesh et al. (2011): we choose galaxies from the F775W UDF image as the sample of sources to be lensed by our simulated cluster mass models. This ensures we have a realistic background field that incorporates the observed distributions of galaxy morphologies, redshifts, luminosities, angular sizes, and ellipticities directly into our simulation. We then simulate the lensed UDF images via ray-tracing, as briefly described in Section 2.3.1. Adopting the thin lens approximation, the lensing can be described by the lens equation,

$$\beta = \theta - \alpha(\theta, z_s), \tag{2.2}$$

where θ is the image position, β is the source position in the source plane, and α is the deflection angle which has a weak dependence of source redshift. Coe et al. (2006) have produced a UDF photometric redshift catalogue and a corresponding segmentation map containing 9821 objects detected above a 8σ level. Based on the redshift catalog, we assign all the UDF sources among

20 redshift bins with bin widths of $\Delta z = 0.3$. In each redshift bin (α is then fixed) we perform the ray-tracing to generate the simulated lensed image and combine each of the simulated lensed objects from all bins into a final image. Finally, we match the noise levels in the simulations to that in the CLASH images.

2.6 Comparison Between Simulated Images and Real Observations

We run the arc-finder on all 640 simulated images. A raw total of 3304 arcs with $l/w \geq 6.5$ and $l \geq 6''$ are detected in 640 simulated realizations. We correct this total number of arcs for elongation bias and incompleteness and obtain 3585 ± 165 arcs, giving a mean of 4 ± 1 arcs per cluster after applying the false positive correction. This value matches the observed lensing efficiency of 4 ± 1 precisely. There is no significant difference between the arc abundance detected in the observations with that detected in the MOKA simulations. Examining the observed and simulated distribution of number of arcs per cluster (Figure 2.14), a Kolmogorov-Smirnov test[2] yields a p-value = 0.92, indicating that the null hypothesis that both distributions are drawn from the same parent distribution cannot be strongly rejected. We further test the lens-

[2]The K-S test performed here use the `ks_2samp` routine from the SciPy package.

ing efficiency as a function of cluster redshift by assigning the observed and simulated samples into two sub-samples by their redshift: $z_{CL} \leq z_{median}$ and $z_{CL} > z_{median}$, where $z_{median} = 0.352$. For each sub-sample, we compare the observed and simulated number distribution (see Figure 2.15) of the lensing efficiency. On average, the higher redshift clusters are slightly more efficient lenses than the lower redshift clusters but the differences are all at marginal statistical significance. The K-S tests indicates that, in both redshift bins, the observed and simulated distributions of the lensing efficiency are consistent with being drawn from a common population (p-values are 0.99 and 0.65 for the lower and higher redshift bins, respectively). We summarize our arc statistics results for the observations and simulations in Table 2.3. The second and third columns in Table 2.3 denote the lensing efficiency of the observed and simulated samples, respectively; the fourth column is the p-value of the K-S test on the observed and simulated distributions.

We now explore the relationships between the lensing efficiency and the cluster's redshift, mass, concentration and effective Einstein radius $\theta_{E,eff} = \sqrt{A/\pi}$ for CLASH and MOKA samples, where A is the area enclosed by the tangential critical curve. Figure 2.16 shows the lensing efficiencies as functions of cluster redshift, mass, central concentration and $\theta_{E,eff}$. Since the CLASH sample does not span very wide range in the cluster redshift, mass and concentration, it is perhaps not surprising that there are no clear correlations between

the lensing efficiency and the redshift, the mass or the concentration for both the CLASH and MOKA samples. However, there is a very significant correlation between the MOKA lensing efficiency and $\theta_{E,eff}$, and the correlation can be described by the following formula:

$$N_{arc} = (0.03 \pm 0.01)\theta_{E,eff}^{1.54\pm0.08}[\text{arcsec}] + (0.81 \pm 0.22), \tag{2.3}$$

as the dashed line in Figure 2.16 [d] shows. The non-zero value of the y-intercept reflects a contribution from false positive detections (consistent with our estimation from simulations) and intrinsic scatter.

Table 2.3: Comparison of CLASH and MOKA's Cluster Lensing Efficiencies.

Redshift Range	CLASH Observations	MOKA Simulations	p-value of K-S test
All Clusters	4 ± 1	4 ± 1	0.92
$z_{CL} \leq z_{median}$	3 ± 1	3 ± 1	0.99
$z_{CL} > z_{median}$	5 ± 1	6 ± 1	0.65

2.7 What is the Dominant Determinant of Cluster Lensing Efficiency?

We now assess the relative importance of the redshift distribution of the lensed sources and the $c - M$ relation of the clusters on the resulting giant arc abundance. We accomplish this by conducting a series of simulations where we

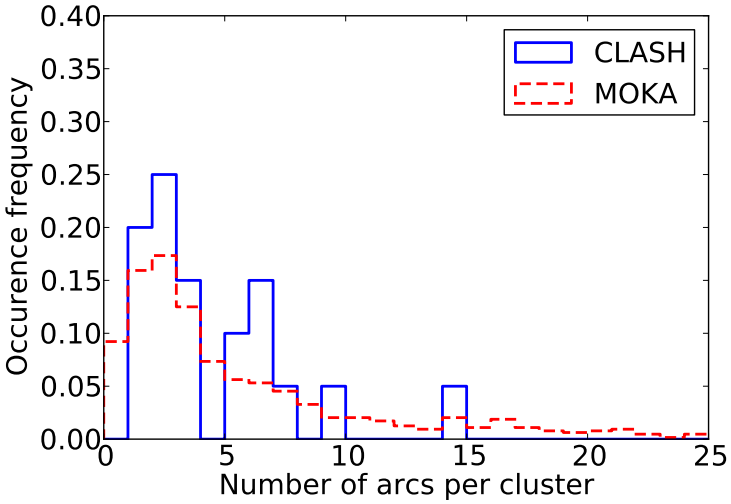
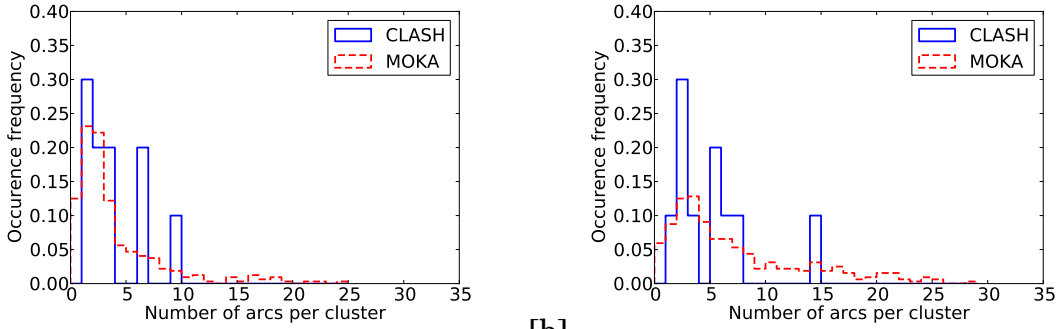


Figure 2.14: The comparison of distribution of arc number per cluster between the X-ray selected CLASH sample and the MOKA simulated sample with same mass and redshift range.



[a] [b]

Figure 2.15: The lensing efficiency as a function of cluster redshift. The 20 X-ray selected CLASH clusters are divided into two sub-samples by their cluster redshift: $z_{CL} \leq z_{median}$ and $z_{CL} > z_{median}$, where $z_{median} = 0.352$. (a) and (b) list the comparison of the number distribution of the sub-samples between the observation and simulation.

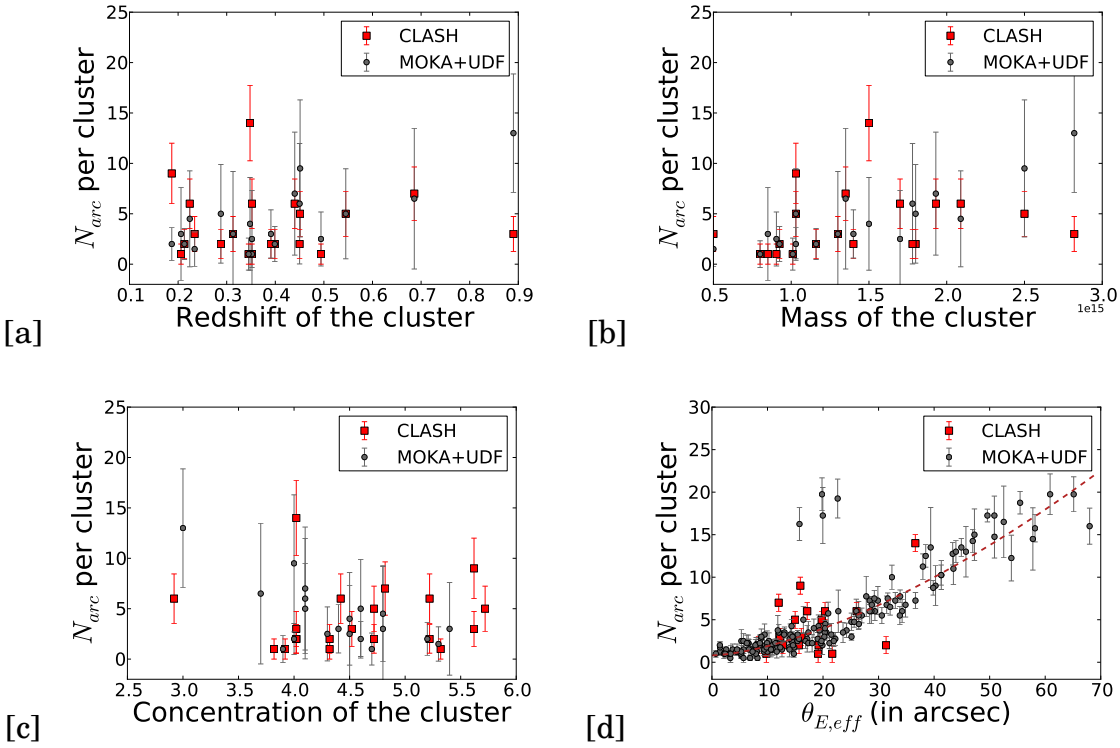


Figure 2.16: Figure (a), (b), (c) show the comparisons of the lensing efficiency between the CLASH and MOKA samples for the corresponding cluster redshift, mass and concentration, respectively. Figure (d) shows the relation between the lensing efficiency and the effective Einstein radius $\theta_{E,eff}$ for all the CLASH and MOKA data points. With the upper left outlier excluded, the dashed line gives the best fitting curve for all the MOKA data points.

alter either the redshift distribution of the background galaxies or the assumed $c - M$ relation. While other effects such as DM substructure, halo triaxiality, and the mass profile of the BCG, may also play a role in determining the distribution of arc number counts, we focus here on studying impact of the redshift distribution and $c - M$ relation as these are potentially the most important effects. As shown below, however, we find that the lensing efficiency of CLASH-like clusters is not very sensitive to the redshift distribution of the background galaxy population so long as there is a significant fraction of the source galaxy population that lies at $z > 1$. We also find that the lensing efficiency is quite sensitive to the DM concentration distribution.

2.7.1 Simulated Lenses, Background Sources all at $z_s = 1$ or $z_s = 2$

We start by testing how the source redshift distribution affects the arc abundance. We use the same 160 simulated MOKA clusters but first set all the UDF source redshifts to $z_s = 1$ and, in a separate realization, then set all source redshifts to $z_s = 2$ to see the impact of a delta function redshift distribution (which is obviously an extreme assumption). We then perform the ray-tracing to create 640 new simulated images for each case. We run arc-finder on these images and detect 1748 and 3764 arcs in total, respectively, when $z_s = 1$ and $z_s = 2$.

After applying statistical corrections, we find lensing efficiencies of 2 ± 1 ($z_s = 1$) and 5 ± 1 ($z_s = 2$). The lensing efficiency decreases by a factor about 2 when the background redshift distribution is a delta function with all sources at $z_s = 1$. However, when putting all sources at $z_s = 2$ one obtains a similar lensing efficiency as that obtained when using realistic UDF redshift distribution. The distributions of arc number per cluster for these 3 cases are shown in Figure 2.17 [a]. K-S tests indicate that the arc number distributions when using the UDF redshift distribution and using a delta function at $z_s = 2$ are consistent (p-value = 0.45). The arc number distribution when assuming a $z_s = 1$ delta function differs significantly from that with UDF redshift distribution or $z_s = 2$ delta function redshift distribution (K-S test p-value = 3.5×10^{-6}).

2.7.2 CLASH Mass Models, UDF Redshift Distribution for the Background Galaxies

Given the CLASH mass models (CLMM), we would like to check if the UDF field is representative as a background source for the simulations. We use the publicly available mass models of 20 CLASH X-ray selected clusters (Zitrin et al., 2015) to lens the UDF source galaxies, and to create 152 simulated images. We detect 656 arcs from these images, corresponding to a lensing efficiency of 3 ± 1. This efficiency differs from that found for the actual CLASH images

(4 ± 1) by 0.7σ. The distributions of arc number per cluster are consistent with one another (see Figure 2.17 [b]). A K-S test gives the p-value = 0.42.

2.7.3 CLASH Mass Models, Background Sources all at $z_s = 1$ or $z_s = 2$

We now assess whether the lensing efficiency is altered significantly when using the CLASH mass models along with delta function redshift distributions. Again, we arbitrarily place all the UDF sources redshift to $z_s = 1$ and $z_s = 2$, and perform ray tracing through 19 CLASH mass model to create 152 new simulated images for each case. We detect a total of 414 and 670 arcs for the $z_s = 1$ and $z_s = 2$ source distributions, respectively. These correspond to lensing efficiencies of 2 ± 1 and 3 ± 1. Similar to that in the MOKA simulations, the lensing efficiency and distribution of arc numbers are similar for simulations with UDF redshift distribution and $z_s = 2$ (p-value = 0.5). Whereas the lensing efficiency for $z_s = 1$ is again about 2 times lower than that with UDF redshift distribution and $z_s = 2$, and the arc number distribution for $z_s = 1$ is also significantly different (K-S p-value = 1.8×10^{-4}). Figure 2.17 [c] shows the distributions of arc number per cluster of the three samples.

2.7.4 Different $c - M$ Relations, UDF Redshift Distribution for the Background Galaxies

Here we show how the arc abundance depends on the cluster $c-M$ relation. Using the UDF redshift distribution, we re-simulate 160 new clusters and simulated images with MOKA by adopting the $c-M$ relation in Neto et al. (2007), instead of Bhattacharya et al. (2013). We detect 230 arcs from 160 realizations using the Neto et al. (2007) $c - M$ relation, which, after corrections, yields a lensing efficiency of 1 ± 1. The lensing efficiency is a factor of 4 lower using the Neto et al. (2007) $c - M$ relation than when we adopt the

2.8 MUSIC Lensing Simulations

Although the lensing efficiency in semi-analytic MOKA simulations is in excellent agreement with that found in the CLASH observations, it is important to make sure this is a robust result. Thus, we study a different suite of simulations to determine the arc abundance using simulated clusters drawn directly from high resolution, hydrodynamical simulations. For this, we use a set of mock clusters taken from the MUSIC-2 N-body/hydrodynamical simulations (Meneghetti et al., 2014). The MUSIC-2 sample (Sembolini et al., 2013; Biffi et al., 2014) consists of a mass limited sample of re-simulated ha-

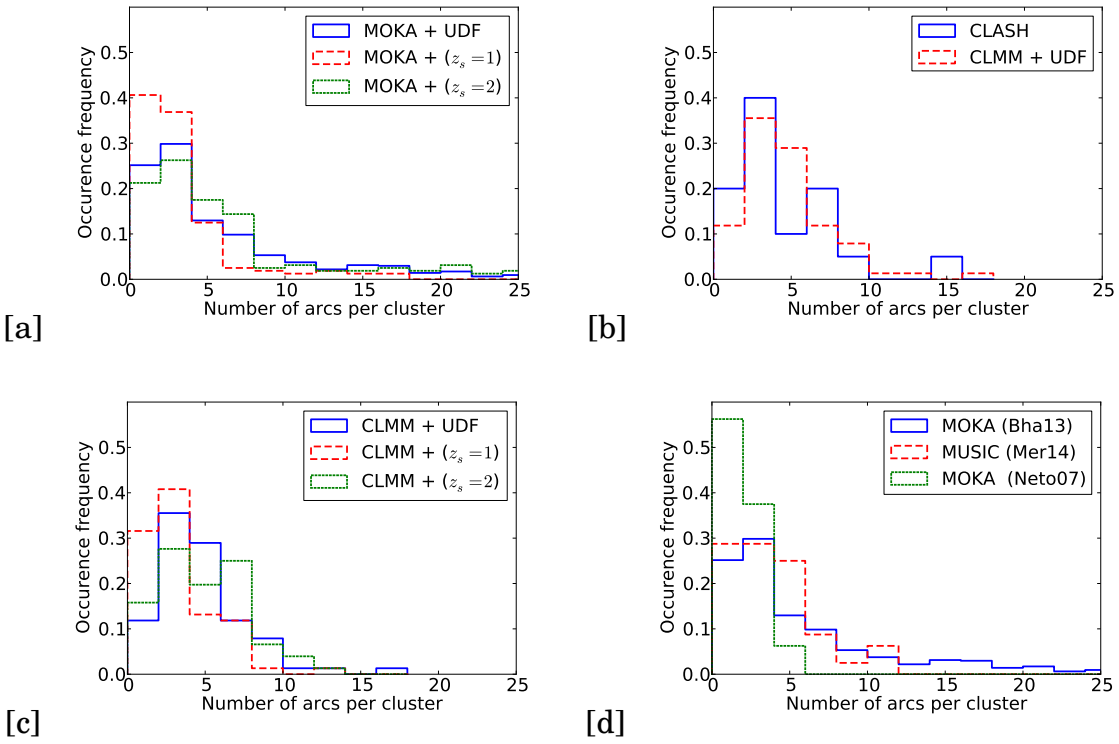


Figure 2.17: Comparison of the distribution of arc number per cluster between different samples. (a) shows the comparison of the arc number distribution between samples with MOKA mass models and different source redshift distributions; (b) shows the comparison of the arc number distribution between the CLASH sample and CLMMs (CLASH mass models) using the UDF redshift distribution; (c) shows the comparison of the arc number distribution between samples with CLMMs and different source redshift distributions; (d) shows the comparison of the arc number distribution between samples with the same source redshift distributions and mass models, but implemented with different $c - M$ relations.

los selected from the MultiDark cosmological simulation. This simulation is dark-matter only and contains 2048^3 particles in a $(1h^{-1}\mathrm{Gpc})^3$ cube, which was performed in 2010 using ART (Kravtsov et al., 1997) at the NASA Ames Research center. All these simulations are accessible from the online MultiDark Database2 . The run was using the best-fitting cosmological parameters to WMPA7+BAO+SNI (Ω_M = 0.27, Ω_b = 0.0469, Ω_Λ = 0.73, σ_8 = 0.82, n = 0.95, h = 0.7). There were 282 cluster-scale halos in the simulation box which are more massive than $10^{15}h^{-1}M_\odot$ at redshift $z = 0$ and are selected to construct our sample. All these massive clusters were re-simulated both with and without radiative physics. The initial conditions for the re-simulations were generated in a finer mesh of size 4096^3, by following the zooming technique described in Klypin et al. (2001). By doing so, the mass resolution of the re-simulated objects corresponds to $m_{DM} = 9.01 \times 10^8 h^{-1}M_\odot$ and to $m_{SPH} = 1.9 \times 10^8 h^{-1}M_\odot$, which was improved by a factor of 8 with respect to the original simulations. The parallel TREEPM+SPH GADGET code (Springel, 2005) was used to run all the re-simulations. Snapshots for 15 different redshifts in the range $0 \leq z \leq 9$ are stored for each re-simulated object. The snapshots which overlap with the redshifts of the CLASH clusters are at $z = 0.250, 0.333, 0.429$ and 0.667.

These re-simulated cluster halos were originally used to estimate the expected concentration-mass ($c-M$) relation for the CLASH cluster sample (Merten et al., 2015; Meneghetti et al., 2014). As in these works, we use the X-ray image

simulator X-MAS (Gardini et al., 2004) to produce simulated Chandra observations of the halos, and use them to further identify objects that match the X-ray morphologies and masses of the X-ray selected CLASH clusters. The $c-M$ relation from our X-ray selected set of simulated clusters agrees with that directly derived from the CLASH data at the 90% confidence level (Merten et al., 2015) and is fully consistent with the stacked weak-lensing signal derived from the ground-based wide-field observations (Umetsu et al., 2014). We perform ray-tracing through these X-ray selected simulated clusters (BCG and radiative physics are not included) to lens the UDF sources and create 100 simulated CLASH images.

2.8.1 Lensing Statistics of MUSIC Simulated Samples and Comparison with Real Observations

We run the arc-finder on the 100 MUSIC simulated images and detect a total of 343 arcs with $l/w \geq 7$ and $l \geq 6''$. We correct the total number of arcs for the elongation bias and incompleteness, yielding a final number of 447 ± 24 arcs, which corresponds to a mean value of 3 ± 1 arcs per cluster after application of the false positive correction. The MUSIC lensing efficiency is fully consistent with the lensing efficiency of the observed CLASH X-ray selected sample (4 ± 1).

Figure 2.18 [a] shows the observed and simulated distributions of arc number per cluster. A K-S test between these two distributions has a p-value = 0.95. We also explored the dependence of the lensing efficiency on the l/w_{min} and l_{min} in the MUSIC simulations (Figure 2.18 [b], 2.18 [c]). The lensing efficiency decreases with increasing l/w_{min} and l_{min} values, which is consistent with the behavior seen in the CLASH observations. We summarize the main arc statistics results of this paper in Table 2.4: the second column in Table 2.4 is the rounded-off value of the mean lensing efficiency (number of arcs per cluster); the third column is the significance of difference in lensing efficiency between the specific simulation sample and that derived for the observed the CLASH X-ray selected sample. As with the MOKA simulations, the MUSIC simulated clusters yield cluster lensing efficiencies that match that seen in the observations when the simulations adopt a $c - M$ relationship and a source redshift distribution that matches the observations.

Table 2.4: Comparison of lensing efficiency between observation and simulation.

	Lensing efficiency	Difference relative to CLASH X-ray selected sample
Observation (X-ray selected sample)	4 ± 1	
Observation (high-magnification sample)	5 ± 1	0.7σ
CLMM + UDF z-distn	3 ± 1	0.7σ
CLMM + ($z_s = 1$)	2 ± 1	1.4σ
CLMM + ($z_s = 2$)	3 ± 1	0.7σ
MOKA + UDF z-distn	4 ± 1	
MOKA + ($z_s = 1$) + (B13) $c - M$	2 ± 1	2.2σ
MOKA + ($z_s = 2$) + (B13) $c - M$	5 ± 1	0.7σ

	Lensing efficiency	Difference relative to CLASH X-ray selected sample
MOKA + UDF z-distn + (N07) $c - M$	1 ± 1	2.2σ
MUSIC + UDF z-distn	3 ± 1	0.7σ

2.9 Discussion

Since the arc statistics was originally proposed as a cosmological probe, many previous studies have investigated the sensitivity of the arc abundance on various cosmological effects. Cosmology enters the strong lensing properties of the galaxy clusters in two ways: first, the arc abundance depends on the angular-diameter distance and volume which are determined by the cosmological expansion; second, the arc abundance depends on the cluster abundance and internal structure which are cosmological sensitive. N-body simulations and semi-analytic approaches have been utilized in earlier studies to explore the sensitivity of arc abundance on σ_8 (Wambsganss et al., 2004; Li et al., 2006; Fedeli et al., 2008) and an increasing function of arc abundance with σ_8 has been observed, though whether the large increments in arc abundance when increasing the σ_8 are quantitatively reliable is not clear; Boldrin et al. (2015) has studied the arc abundance dependence on σ_8 and Ω_m for a given survey area. They use MOKA to generate mock clusters with different mass and redshift and populate them into the light cones spanned by the survey region.

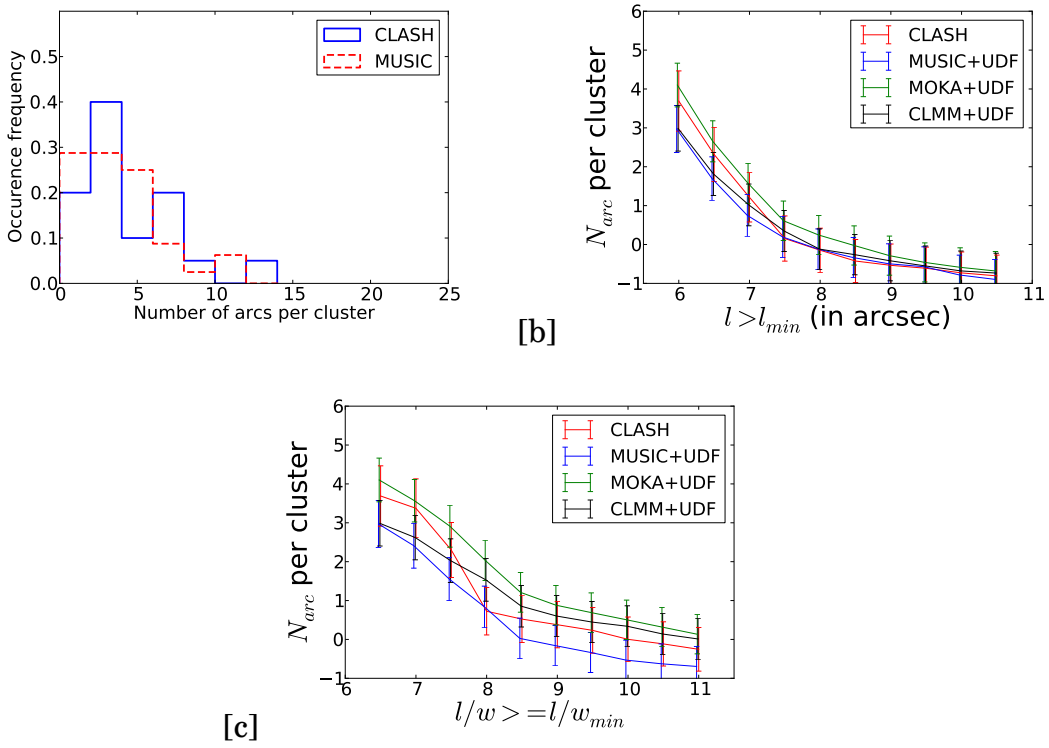


Figure 2.18: (a) shows the comparison of the number distribution between CLASH sample and MUSIC sample; (b) shows the lensing efficiency as a function of l_{min} for arcs with $l/w \geq 7$ for different samples; (c) shows the lensing efficiency as a function of l/w_{min} for arcs with $l \geq 6''$ for different samples;

They identify the increasing functions of arc abundance with both parameters. The arc abundance seems more sensitive to σ_8 than Ω_m, because σ_8 has an effects on the cluster formation time, which in turn affects the cluster internal lensing properties such as triaxiality and concentration. However, the degeneracy between two parameters for the arc abundance limit its ability to distinguish different cosmologies. The arc abundance sensitivity on various dark energy models has also been studied, which includes a constant equation-of-state parameter $w \neq 1$ (Bartelmann et al., 2003) and time varying w parameter (Meneghetti et al., 2005), the arc abundance could change by at most a factor of a even with substantial change in w. Jullo et al. (2010) and D'Aloisio & Natarajan (2011b) studied how cosmology affects the arc statistics through geometry effects. They found the expansion function thus the cosmological models can be constrained from the ratio of the lensing efficiencies at different redshift. To achieve competitive results, however, the mass distribution of the clusters must be determined with very high precision, and a sample of about ten clusters containing about 20 arc families each are needed. Moreover, the arc statistics could even change by 30% with different non-Gaussianity parameters based on theoretical framework D'Aloisio & Natarajan (2011a). Therefore, these studies may indicate that, amongst all the cosmological parameters, the arc abundance seems to be most sensitive to σ_8. Interestingly, most of the simulations in early arc statistics works have adopted a typically higher σ_8

value $\sim 0.9 - 0.95$ ($\sigma_8 = 1.12$ was adopted in B98) than the current consensus from WMAP7 and PLANCK, which could have made the discrepancy between the simulations and the observations even larger. It implies that at least the deficit of cluster abundance under different cosmologies might not be the main solution to "arc statistics problem" in the first place. Since the dependence of the cluster internal lensing properties on σ_8 is still not well known, we simply adopt the value of $\sigma_8 = 0.83$ along with other cosmological parameters from the Planck results (Planck Collaboration et al., 2014). We believe that our conclusion would not change significantly unless there is large revision in the Planck value for σ_8.

As one of the promising candidate solutions to the arc statistics problem, the impact of source redshift distribution on the arc abundance has been emphasized by many previous studies. Wambsganss et al. (2004) studied the magnification probability for light rays propagating cross a cosmological scale and found that the probability of high magnification events highly depends on the source redshift. They concluded that the arc abundance should have a steep increase with source redshift because the number of halos suitable for strong lensing increases exponentially with redshift, and they suggest this as the solution for the arc statistics problem. Bayliss et al. (2011); Bayliss (2012) has established a large sample of arcs (105) from the Sloan Giant Arcs Survey (SGAS) and from the Second Red Sequence Cluster Survey (RCS2) and study the red-

shift distribution of the arcs. They find that arcs with $g \leq 24$ have a median redshift of $z_s \sim 2$. Bayliss (2012) claim the arc statistics problem can be solved by adopting their measured redshift distribution and using the scaling of the optical depth given in Wambsganss et al. (2004). However, Li et al. (2005) and Fedeli et al. (2006) show that the scaling of the optical depth is very different from what assumed by Wambsganss et al. (2004) and that the Wambsganss et al. (2004) assumption that the magnification is a good measure for the l/w ratio of an arc is not justified in detail. Furthermore, it is unclear if the arc identification used by Bayliss (2012) (e.g. by curvature radius of arcs and by visual inspection) might bias the selection in favor of luminous and highly curved arcs. If so, the corresponding arc abundance and redshift distribution could also possibly be biased.

Our results show that the simulations performed either with a UDF redshift distribution or a delta function redshift distribution at $z_s = 2$ give very similar arc abundances. When we change the redshift distribution of the background sources to a delta function at $z_s = 1$ the arc abundance drops by a factor of 2 rather than the order of magnitude change in the arc abundance noted in some previous studies (e.g., Wambsganss et al. (2004)). The factor of 2 change is consistent with Horesh et al. (2005, 2011), who also used UDF images as background sources to perform the ray-tracing. Horesh et al. (2005) used the same simulated clusters at $z_c = 0.2$ as used in B98 to lens the UDF sources

and found an arc abundance that was 3 times higher than that in B98. They attributed this over-abundance to the use of a source number density that was 3.2 times higher than that in B98. They found that changing the source redshift distribution from a delta function at $z_s = 1$ to a realistic UDF distribution results in only a small change in the final arc abundance. These results suggest that the redshift distribution does not have a major impact on the final arc abundance unless one selects a distribution that significantly underpopulates galaxies in the $z > 1$ range.

The MUSIC simulated halos do not have BCGs at the center and do not implement complex gaseous physics. However, Killedar et al. (2012) has compared the arc production efficiency of the adiabatic simulations with some more sophisticated simulations which include the effects such as gas cooling, star formation, feedback from AGNs and SN+galactic winds, etc. The comparable results indicates that the implementation of baryonic physics will probably not lead to a significant change in the arc abundance derived from simulations without such processes.

Previous studies have already revealed the correlation between the lensing cross section and the Einstein radius, $\theta_{E,eff}$, from N-body simulation (Meneghetti et al., 2011) and the semi-analytic calculations (Redlich et al., 2012). Our study confirms this correlation as reflected by the dependence of the number of arcs per cluster on $\theta_{E,eff}$, as shown in Figure 2.16 [d] shows. The relation between

the MOKA cluster lensing efficiency and $\theta_{E,eff}$ in our study is well fit by a linear relation in log-log plane with a slope of 1.54 ± 0.08, which is flatter than the slope 1.79 ± 0.04 in Meneghetti et al. (2011) and 2.4 ± 0.04 in Redlich et al. (2012). The detection of this correlation in our MOKA simulations is due to the relatively large size of the MOKA cluster sample (640 simulated clusters), whereas the CLASH sample is too small to robustly unveil this correlation. For the arc abundance ~ 5 per cluster, the fractional error for an ensemble of 32 realizations is $1/\sqrt{5 \times 32} \sim 8\%$. Therefore, to measure the correlation observationally to 10% and to detect a $\sim 15\%$ deviation from such correlation, we need $\frac{\times(1/0.1)^2}{5} = 20$ clusters in each mass bin, and we probably need a cluster sample with size ∼200 if 10 different mass bins are expected.

We are able to identify the relative significance of several key physical effects which contribute to the arc abundance enhancement. As Table 2.4 shows, varying the source redshift distribution leads to, at most, a factor of 2 variation in the arc abundance. Variation of the $c - M$ relations will affect the matter distribution of the inner cluster core and, hence, lead to variations in the arc abundance. Using several recent estimates of the $c - M$ relation (Neto et al., 2007; Bhattacharya et al., 2013) results in variations of the arc abundance by up to a factor of $\sim 4 - 5$. Using the most recent estimates of the $c - M$ relation in simulations appears to produce excellent agreement with the observed arc abundance. However, quantities such as mass and concentration alone are not

sufficient to reflect the likely complex dependencies of the arc abundance on various effects. As shown in Figure 2.16 [b] and 2.16 [c], the arc abundance fails to exhibit a strong dependence on either the concentration or the cluster mass alone, for both the CLASH and MOKA samples. By contrast, the effective Einstein radius, $\theta_{E,eff}$, is a good indicator of the lensing efficiency.

Given our results, even without fully understanding the cosmological dependence of the arc abundance, we could still conclude that the initial "arc statistics problem" appears to have been largely due to inadequate modeling of the mass distributions of the clusters and, secondarily, due to inadequate modeling of the background source number density and redshift distribution. In addition, the previous use of mostly visual identification of arcs may have resulted in an inadequate modeling of the false positive contamination rate and completeness corrections. We can divide the contributions from different physical effects on cluster lensing efficiency into three general categories: the cluster abundance, the background source redshift distribution, and the individual cluster lensing cross section. Our study would suggest that the lensing efficiency is more strongly dependent on the individual cluster lensing cross sections than on the source redshift distribution. However, different cosmology could alter both the cluster abundance and the individual cluster lensing cross sections and the relative significance of such factors has not been explored in this study given the small cluster sample size. Future large cluster surveys

(e.g., DES, LSST, Euclid, WFIRST) will definitely help to answer this question. We suspect that two other related problems in lensing, the over-concentration problem and Einstein radii problem, where it has been found that some real clusters at intermediate redshift have denser cores than clusters of similar mass produced in simulations (Broadhurst & Barkana, 2008; Oguri & Blandford, 2009; Richard et al., 2010; Sereno et al., 2010; Merten et al., 2015) and where some real clusters have larger Einstein radii than expected in standard ΛCDM cosmology, may well be due to a combination of insufficiently accurate cluster simulations and observational sample selection effects.

2.10 Summary

We have carried out an observational and theoretical study of the arc statistics problem in clusters of galaxies. We have devised an automated arc-finder to efficiently and objectively detect arcs. We test our arc-finder using a large number of simulated cluster images and have quantified the incompleteness and false positive rate in arc detection. We also investigate how image noise affects the shape determination of the arcs and statistically correct for the observed elongation bias. We run our arc-finding algorithm on 20 X-ray selected CLASH clusters and 5 high-magnification CLASH clusters. After correcting for arc shape elongation bias, incompleteness and false positive rate we find a

large arc ($l/w > 6.5$ and $l \geq 6''$) lensing efficiency of 4 ± 1 arcs per cluster and 5 ± 1 arcs per cluster, respectively, for the X-ray selected and high-magnification selected CLASH samples.

We simulate mock clusters using both the MOKA semi-analytic cluster generator and the MUSIC-2 N-body results. In both cases, we focus on simulated clusters that have the same mass and redshift range as the CLASH clusters. For the MOKA simulations, we use ray-tracing to create 640 simulated cluster realizations with the F775W UDF image as the background source. For the simulations extracted from the high resolution, hydrodynamical simulations (MUSIC), we identify halos that, in addition to having similar redshifts and Virial masses as the CLASH clusters, are also selected to have similar X-ray morphologies as the CLASH clusters. We find a lensing efficiency of 4 ± 1 arcs per cluster in the MOKA sims and 3 ± 1 arcs per cluster in the MUSIC sims. These lensing efficiencies both match the observed lensing efficiency of 4 ± 1 arcs per cluster. We also study the arc abundance dependence on the cluster redshift by splitting the sample into two bins divided at the median sample redshift of $z_{median} = 0.352$ and find no significant differences in either the overall lensing efficiency and arc redshift distributions. The dependence of the MOKA and MUSIC lensing efficiencies on l_{min} and l/w_{min} also match that seen in the observed CLASH ones.

For the future, the relative short running time (less than 5 minutes for

images with 3000×3000 pixels) of our arc-finder allows us to perform large-scale "blind" searches for giant arcs in various other surveys, especially those with moderately high-angular resolution such as WFIRST and Euclid. Moreover, continued study of the correlation between the arc abundance and the $\theta_{E,eff}$ should be conducted to assess just the reliability of using arc abundance (which is an observable) as a predictor of $\theta_{E,eff}$.

Chapter 3

The CLASH High Redshift Lensed Galaxy Sample

3.1 Introduction to Study of Highly Lensed Galaxies

Two decades of observations by Hubble Space Telescope (HST), IFUs on VLT as well as GALEX, Keck, etc, have unveiled the rich physical properties of high-redshift star-forming galaxies. The cosmic star formation rate density peaks between $z = 1.5$ and $z = 2.5$ (Madau & Dickinson, 2014) and marks a critical period for galaxy assembly. The irregular morphologies of the high-redshift star-forming galaxies (SFGs) has been widely identified in the rest-frame UV

images (Windhorst et al., 1994; Abraham et al., 1996) and rest-frame optical images (Elmegreen et al., 2009; Forster Schreiber et al., 2011). These SFGs are compact and their size (1-5 kpc) is typically $3-5$ times smaller than their low-redshift counterparts (Trujillo et al., 2007; Buitrago et al., 2008; Szomoru et al., 2011). How these star-forming active, irregular and compact galaxies evolve to star-forming ceased, regular and more extended local galaxies is still an open question. In the classic picture of galaxy formation, such compact and irregular systems with high star formation rate, can be explained by some violent dynamical process such as major merger (Hopkins et al., 2006, 2008). However, evidence has been accumulating over the past decades that not all the high-z galaxies undergo major merging. Studies of galaxy pairs counting (Bundy et al., 2009; de Ravel et al., 2009; Williams et al., 2011) and the disturbed morphologies of galaxies out to $z = 2$ (Lotz et al., 2008, 2011; Conselice et al., 2009) have suggested that the pair/merging percentage ranges from a few to at most 25%. The kinematics studies also have shown that majority of SFGs is dominated by ordered disk rotation (Shapiro et al., 2008; Forster Schreiber et al., 2009; Genzel et al., 2011; Epinat et al., 2012), which is different from the dispersion-dominated system expected in frequent merger scenario. Moreover, the "main sequence of star formation" – a well-defined relation between the ongoing star formation rate and the amount of the assembled stellar mass has been observed up to $z_s \sim 2$ (Noeske et al., 2007; Daddi et al., 2007). The

tightness of the star-forming main sequence supports the gradual evolution of star-forming history (Robaina et al., 2009; Renzini et al., 2009; Bouche et al., 2010; Wuyts et al., 2011; Rodighiero et al., 2011), instead of a stochastic growth trajectory predicted by major mergers.

The typical star formation rate (SFR) of SFGs at $z \sim 2$ is about 100 times higher than that in a low-redshift disc like the Milky Way. If not mergers, other mechanisms must play a dominant role in fueling the high level of star formation activity. Cosmological simulations indicate that at that epoch, filamentary cold gas streams can efficiently deposit new fuel for continuous star formation (Keres et al., 2005; Dekel & Birnboim, 2008; Dekel et al., 2009a). In this picture, several massive star-forming clumps will form within the continuously replenished gas-rich disks due to the gravitational instabilities. The existence of these massive clumps not only provides a natural explanation of the irregular morphology and high SFRs; but also provides an interesting clue to the build-up of the central bulge: though the longevity of the clumps is still a hot topic of debate, if these clumps can survive for sufficiently long time, the global gravitational torque exerted by host galaxies would drive them migrate towards the center of galaxy, where they end up to coalesce into a bulge (this "prediction" is hereafter referred to as the "theory of clump migration" in subsequent sections of the thesis), which have been widely identified in numerical simulations (Noguchi, 1999; Immeli et al., 2004a,b; Dekel et al., 2009b; Cev-

erino et al., 2010; Bournaud et al., 2011).

Therefore, the massive clumps, which may be the building blocks of the clumpy structure in high-z SFGs, become a key ingredient to not only understand how the transition from the clumpy and irregular SFGs at high-z to the low-z galaxies with regular morphology occurs, but also how and when the galactic bulge is assembled. To investigate the formation and evolution of the massive clumps, very detailed information about the integrated physical properties such as SFRs, stellar mass, metallicity and morphology as well as the spatial-resolved physical properties are needed. Since the typical size of the clumps is at sub-kpc scale (Elmegreen et al., 2007; Forster Schreiber et al., 2011), the current observational studies of these individual star-forming regions are limited by the available spatial resolution. At FWHM $\sim$ 0.1", which corresponds to roughly 1 kpc at $z = 2$, both HST imaging and adaptive optics (AO)-assisted integral field spectroscopy (IFS) at 8-10 m class telescopes have very limited ability to resolve the clumps at sub-kpc scale at this redshift. Before the advent of the next generation space telescope with higher spatial resolution, the strong gravitational lensing of high-z galaxies by the foreground galaxy clusters provides us a unique opportunity to explore the sub-galactic scale structures at a unprecedented spatial resolution, because the strong lensing effect can increase the S/N and apparent size of background galaxies by a factor of a few. Studies of lensed galaxies at high-z have already provided us

some detailed view of high-z galaxies for kinematics up to $z \sim 5$ (Nesvadba et al., 2006; Swinbank et al., 2009; Jones et al., 2010), as well as the individual star-forming regions (Livermore et al., 2012, 2015). In some extreme systems with very high magnification, the spatial resolution can reach up a surprising ~ 10 pc (Rigby et al., 2017; Vanzella et al., 2017), the scale of star cluster.

In this study, we extend the previous pioneering work on the strongly lensed galaxies, based on a larger sample of 106 detected lensed galaxies spanning the redshift range $0.5 < z_s < 6$. The high spatial resolution achieved by strong lensing combined with the availability of multi-wavelength image data enable us to measure the physical properties such as photometric redshift, stellar mass, star formation rate, specific star formation rate, color, age, metallicity, etc, of the lensed galaxies as well as the resolved sub-galactic structures. The high spatial resolution also allows us to study the spatial distribution of the sub-galactic structure, which is crucial to assess the plausibility of the clump migration scenario. In Chapter 3, 4 and 5, we will address the following issues:

1. What is the typical morphology of the high-z galaxies?

2. What are the physical properties of the sub-galactic structure (clumps) in high-z galaxies?

3. What are the luminosity and mass functions of the sub-galactic structure (clumps)?

4. How are sub-galactic structures (clumps) spatially distributed and how

does that distribution evolve with redshift?

5. Is the evolution of the sub-galactic structure (clumps) correlated with that of the host galaxies?

6. Why is the typical mass of the smallest, current-epoch dwarf galaxies comparable to the typical mass of the hi-z clumps? Is this a coincidence or a clue about the origin of dwarf galaxies?

In chapter 3, I describe the selection of the lensed galaxy sample, the measurements of their integrated and spatially-resolved physical properties as well as their morphological classification. In Chapter 4, I describe the study of the sub-galactic structure (clumps) in these lensed galaxies, which includes the identification of the clumps, the measurements of their physical properties, the determination of their rest-frame UV luminosity function and stellar mass function and comparison with other stellar system with different mass and size. In Chapter 5, I describe the spatial distribution of the sub-galactic structure (clumps) and their evolution. I also discuss the compatibility of CLASH data with the clump migration scenario, what we can learn from clump migration and the implications for galaxy evolution. Throughout this study, we adopt a ΛCDM cosmology with parameters $\Omega_m = 0.31$, $\Omega_\Lambda = 0.69$, $\sigma_8 = 0.82$, $H_0 = 100h$ km s^{-1} Mpc^{-1}, and $h = 0.68$ (Planck Collaboration et al., 2014).

3.2 Arc Selection Criteria

The CLASH observations for each cluster consist of 16 broadband images (spanning the range 0.23μm - 1.6μm) using the $WFC3/UVIS$, $WFC3/IR$, and ACS/WFC instruments onboard HST. We run our arc-finder that was described in chapter 2 on the synthetic ($ACS + WFC3/IR$) detection image created for each cluster. We select arc mainly based on two quantities: the integrated signal-to-noise ratio of lensed galaxies S/N and the axis ratio of lensed galaxies l/w. The integrated S/N is defined as:

$$\mathrm{S/N} = \sum \mathrm{I_i} / \sqrt{\sum (\mathrm{VAR_i} + \mathrm{I_i})}, \tag{3.1}$$

where I_i is the intensity of pixel i that belongs to the arc and $\mathrm{VAR_j}$ is the corresponding variance of the noise at pixel i. Since the brighter and more extended arcs serve our purpose of studying overall and spatially-resolved physical properties, we selected 106 giant arcs from 20 X-ray selected CLASH clusters with $S/N \geq 5$, $l/w > 6.5$ and the angular length $l \geq 3$" in HST F160W band as our sample for subsequent studies.

3.3 Arc Photometry

3.3.1 PSF Matching

Because of the broad filter coverage of CLASH observations, the difference in PSF FWHM in different filters needs to be considered. e.g. the PSF FWHM in HST F160W filter is about 4 times larger than that in F435W. Therefore, within the same aperture with small radius, a point source would contain more of its flux in UV or optical filters than in IR filters. We correct for this by degrading the PSF of each frame to that of the F160W filter. We use a Gaussian kernel to fit the PSF function of each frame. The images were then convolved with the appropriate Gaussian kernel (quadratic difference of FWHM of F160W's Gaussian kernel to that of the corresponding frame) to create the PSF-matched images.

3.3.2 Background Subtraction and Photometric Error Estimation

The PSF-matched photometry was then measured in the image plane. The segmentation map (aperture) for each lensed galaxy was generated by the arc-finder. The local background and noise levels were measured as follows: first, we crop out the main body of the lensed galaxies from their segmentation maps,

and iteratively clip out the brightest pixels to decrease the contaminations from the nearby bright sources, until the intensity of all the remaining pixels converge to 2σ about its median value. We then evenly divide the image into 5×5 sub-regions. In each sub-region, we repeat the above pixel-clipping process until convergence is achieved, and calculate its own median value and standard deviation of the pixel intensity, and adopt them as the values of background level and rms of the central pixel in that sub-region. We then use bi-cubic interpolation to interpolate the 5×5 grid of background level and rms values to the full background level and rms images. The total noise level equals the square root of the quadratic sum of the photon noise from the source and the rms value for all the pixels within the aperture.

3.4 Arc Photometric Redshift Estimation

The photometric redshift of the lensed background galaxies were estimated by the Bayesian-based BPZ package (BPZ; Benítez (2000, 2004); Coe et al. (2006)). Spectral energy distribution (SED) templates are redshifted and fit to the observed photometry. The BPZ code adopts a prior that the empirical likelihood of redshift is a function of both galaxy magnitude and galaxy morphological type (e.g., bright and/or elliptical galaxies are rare at high redshift). We used 11 SED templates originally from PEGASE (Fioc & Rocca-Volmerange,

1997) that have been recalibrated based on photometry and spectroscopic redshifts of galaxies in the FIREWORKS catalog (Wuyts et al., 2008). Figure 3.1 shows the comparison of spectroscopic redshift photometric redshift for CLASH spec-z sample from Jouvel et al. (2014). The estimated uncertainty of the photometric redshift is $\delta z \sim 3\%(1+z)$ (Jouvel et al., 2014). We assess how the uncertainty of the measured photometric redshift affects our results in Section 4.8.2 and 4.9.2.

3.5 Overall Magnification Factor Estimation

The observed photometry in the image plane has been amplified by the strong lensing effect. Thus to obtain the true photometry, we need to correct for the magnification factor. We calculate the magnification factor $A(z)$ based on the redshift dependent convergence $\kappa(z)$ and shear $\gamma(z)$ maps from the CLASH lensing models (Zitrin et al., 2015). The formula is given as:

$$A(z) = \frac{1}{|(1-\kappa(z))^2 - \gamma(z)^2|} \tag{3.2}$$

In general, one can calculate the median value of the magnification map within the apertures as the overall magnification factor. However, the spatial

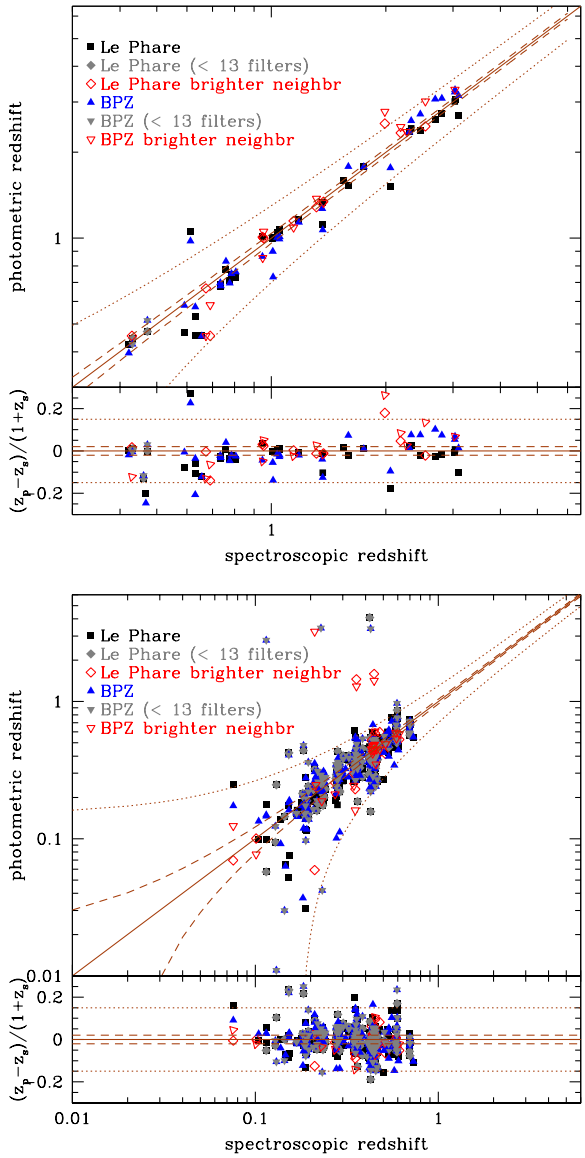


Figure 3.1: The spectroscopic redshift VS the photometric redshift for the CALSH spec-z sample. The black squares and blue triangles are the Le Phare and BPZ photo-z for galaxies observed in at least 7 bands with secure spec-z. The grey triangles and grey diamonds are respectively BPZ and Le Phare secure redshifts observed in less than 13 filters. The empty red triangles and diamonds are the spec-z with a brighter neighbor within 3" of the closest match to the spec-z. The top figure corresponds to galaxy lensed images behind the cluster at a redshift of $z_s > z_{cl} + 0.2$. The bottom figure corresponds to galaxy cluster members and foreground galaxies at $z_s \leq z_{cl} + 0.2$. Figure from Jouvel et al. (2014)

distribution of the magnification is highly uneven. In the case of giant lensed arcs, the tangential direction is more magnified than the radial direction. So the overall magnification factor should be a integrated property which takes the spatial gradient into account. In this study, we calculate the H band flux-weighted magnification factor and adopt this as an overall magnification factor for photometry correction in all the filters.

3.6 SED Fitting Procedure

3.6.1 General Methodology of SED Fitting

Given the estimated photometric redshift and magnification corrected photometry, now we turn to the modeling of the SED fitting to extract the physical properties of the lensed source galaxies. In this study, we make use of state-of-art software iSEDFIT (Moustakas et al. , 2013) to fit the SEDs of the lensed galaxies. There are several parameter settings before running iSEDfit: first, we need to use the stellar population synthesis (SPS) models to specify an input spectral evolution $S(\lambda, t, Z)$ of a simple stellar population (SSP) at a given metallicity Z. An SSP is a stellar population that forms in an instantaneous burst of star formation which then evolves with time t passively. Such SSP includes three basic ingredients: (1) an assumed initial mass func-

tion (IMF), which specifies the initial stellar mass distribution among different stellar populations; (2) stellar evolution calculation that covers the full range of the IMF; (3) a stellar library that provides emergent spectra at each position in the Hertsprung-Russell (HR) diagram. Second, a parametric form $\psi(t)$ of the star formation history (SFH) needs to be specified. For example, one can choose the exponentially declining SFH (simple τ model; Sandage 1986):

$$\psi(t) = \frac{M_{tot}}{\tau} \exp(-t/\tau); \tag{3.3}$$

where t is the age of the stellar population since the onset of the burst, τ is the characteristic timescale for the star formation and M_{tot} is the normalization which is defined to be $M_{tot} = 1M_{\odot}$. Third, a wavelength dependent dust attenuation or extinction law $A(\lambda, t)$ needs to be specified.

Given the input $S(\lambda, t, Z)$, iSEDfit computes the evolution $C(\lambda, t, Z)$ of the integrated SED of a composite stellar population by following the integral:

$$C(\lambda, t, Z) = \int_0^t \psi(t - t')S(\lambda, t', Z)10^{-0.4A(\lambda,t')}dt' \tag{3.4}$$

To better explore the multi-dimensional parameter space, Markov Chain Monte Carlo (MCMC) algorithms are implemented in iSEDfit. A grid of model parameters Q is generated by randomly sampling from a bounded parameter space setting, where Q represents parameters such as stellar mass, SFR, age,

metallicity, etc. For a given Q, the broadband flux F_i at redshift z in filter i, and the corresponding uncertainties σ_i, the goodness-of-fit statistics χ^2 can be evaluated by:

$$\chi^2(F_i, z|Q) = \Sigma_{i=1}^{N} \frac{[F_i - AC_i(Q, z)]^2}{\sigma_i^2}, \tag{3.5}$$

where A is a normalization factor, and $C_i(Q, z)$ are the synthetic broadband fluxes of each model SED given the redshift and parameter Q. After assuming a prior probability $p(Q)$, then one can assesses the posterior probability distribution (PDF) of the model parameter $p(Q|F_i, z)$ by the Bayes' theorem:

$$p(Q|F_i, z) = p(Q) \times p(F_i, z|Q), \tag{3.6}$$

where $p(F_i, z|Q)$ is the likelihood of $L \propto \exp[-\chi^2(\mathrm{F_i, z|Q})/2]$. The value and uncertainty of the model parameter Q are given by the median and 1/4 of the 2.3-97.7 percentile range of the posterior distribution. For a detailed discussion of iSEDFIT, we refer the reader to Moustakas et al. (2013).

3.6.2 Parameter Setting of SED Fitting

In this study, we adopt the initial mass function (IMF) of Chabrier (Chabrier, 2003) and an updated version of Bruzual & Charlot (2003) (BC03), which includes a new treatment on the contribution of thermally pulsing asymptotic giant branch stars, as our stellar population library. BC03 generates a grid of

spectra, varying the stellar population metallicity, age and star formation history. BC03 model the metallicities run from 0.04 (Z = 0.0008) to 2.5 (Z = 0.05) times the solar metallicity, and the ages range from 0.01 Gyr to the age of the Universe at corresponding redshift. We adopt the exponentially declining SFH of the form $\mathrm{SFR(t)} \sim \exp(\mathrm{t}/\tau)$, with e-folding time τ spanning 0.01 - 10 Gyr. We use a Calzetti extinction law (Calzetti et al., 2000) with extinctions in the range $0 < A_V < 2.5$. For each galaxy, we adopt a uniform prior, and generate 10000 model spectra to explore the full posterior probability distribution of the physical parameters.

3.6.3 Estimation of Integrated Stellar Mass and Star Formation Rate

Figure 3.2 show the distribution of the stellar mass and star formation rate (SFR) of CLASH lensed galaxies. We also compare the CLASH distributions with those from CANDELS wide field survey UDS (Santini et al., 2015). For consistency, we select one set of their measured stellar mass and SFR data which are based on similar parameter setting (their method $2a_\tau$) with ours: BC03 stellar templates, Chabrier IMF, simple exponentially declined SFH, τ ranging from 0.3 to 1 and a Calzetti extinction law. For stellar mass, the median value of CLASH sample is $10^{9.4}\mathrm{M}_\odot$. Comparing with CANDELS sample,

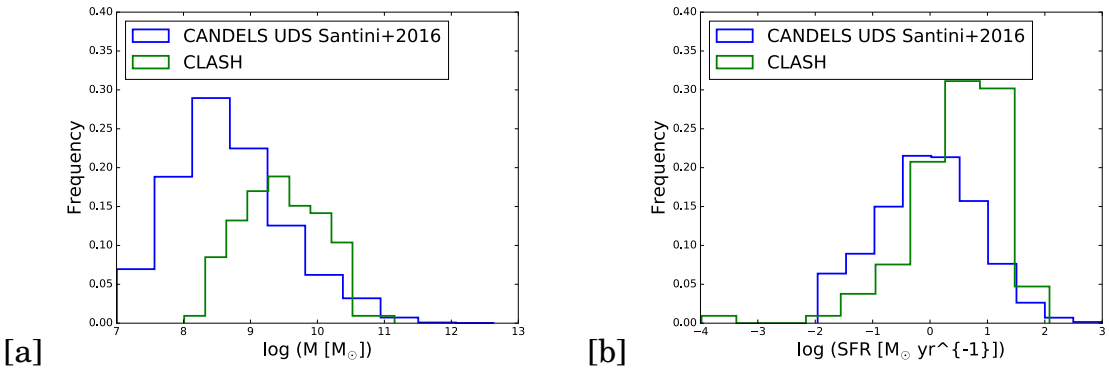


Figure 3.2: The distributions of stellar mass and SFR of CLASH lensed galaxies and CANDELS UDS galaxies.

the distribution has a clear deficit at low mass end $\mathrm{M} < 10^9\mathrm{M}_\odot$, whereas it has a slight excess at high mass end though the slope is similar. This is likely due to the combinational effects of incompleteness and our selection criteria: the higher the magnification is, the smaller the projected sampling area in the source plane is. We select the lensed galaxies based on their axis ratio or magnification, such selection criteria would preferentially samples more abundant thus less massive background objects. For SFR, the median value of CLASH sample is 4 $\mathrm{M}_\odot\mathrm{yr}^{-1}$ and the distribution of SFR is more skewed toward the higher SFR end and has a slightly deficit at low SFR end, which indicates that the lensing selected sample appears to have preference for the more star-forming systems. Overall, the strong lensing effect tends to sample relatively small galaxies with mild SFR, comparing with violent starburst galaxies.

3.7 Sampling Volume Test

Measuring the sampling volume as a function of redshift and lensing magnification is a key component to correcting the raw distribution function of any physical properties (e.g. rest-frame UV luminosity function, stellar mass function) to a form that can be properly interpreted in the context of galaxy evolution. In this section, we make use of the CLASH mass models to study the lensing sampling volume at different redshifts and magnifications.

We first compute the total sampling volume at each redshift based on the lensing models. The calculations proceed as follows: at each redshift bin ($\Delta z = 0.5$), we locate regions in the image plane with magnification factors ranging from 2 to 40. We de-lens the regions within these magnification ranges to the source plane and calculate the total projected surface area. The total surface area is then used to calculate the total solid angle and thus the corresponding volume at each redshift. Figure 3.3 shows the average sampling volume as a function of magnification at each redshift, and the total volume as a function of redshift. As expected, the sampling volume decreases as magnification increases. The total sampling volume appears not to change much over all the redshift range (a factor of 2-3 lower at $z_s \sim 1$), and remains roughly at $\sim 10^4$ Mpc^{-1}. We compute the rest-frame UV luminosity function and stellar mass function in Section 4.8 and 4.9 based on the redshift and magnification

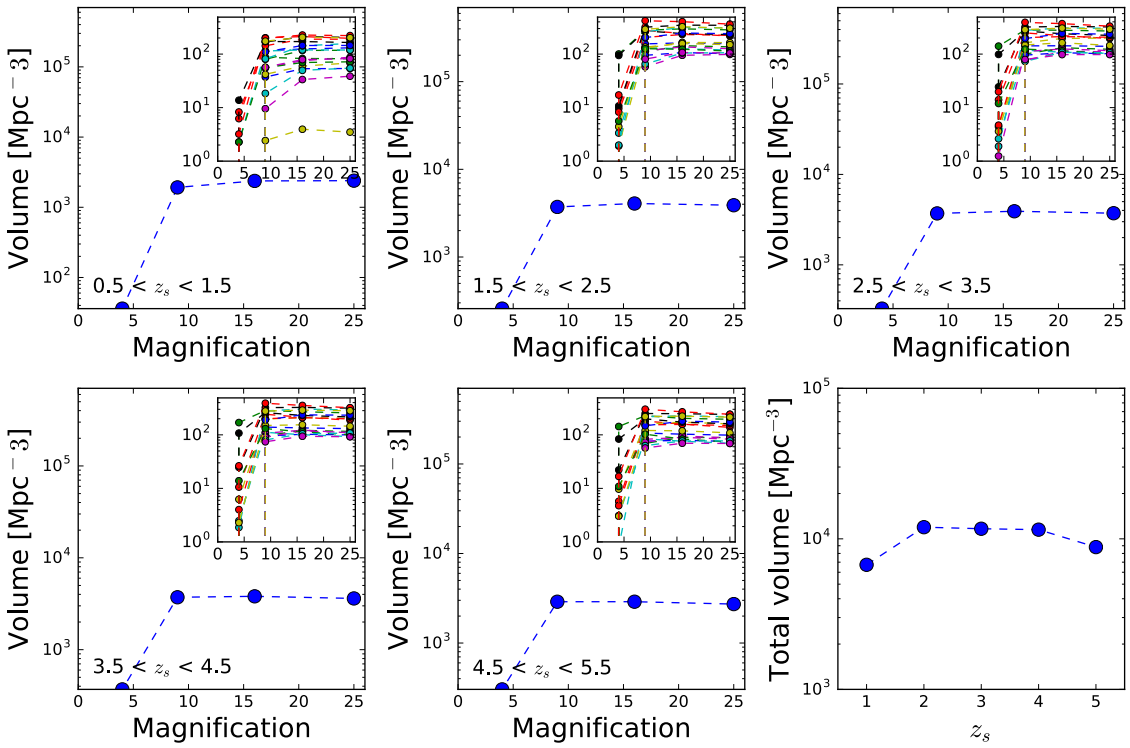


Figure 3.3: The sampling volume as a function of magnification at 5 redshift bins and the total sampling volume as a function of redshift. The magnification dependent sampling volumes for each CLASH cluster are shown in the upper right of each figure.

dependent sampling volumes.

3.8 Spatially-Resolved SED Fitting

Aside from the integrated physical properties, the spatially resolved physical properties are essential if we want to explore the sub-galactic structures and their spatial distributions. To create spatially resolved maps, a relatively high average pixel S/N ratio level (say, $S/N \geq 5$ per pixel) is needed, oth-

erwise the pixel binning would tend to erase some substructure information. Fortunately, the lensing power of our cluster sample allows us to achieve the required S/N level for all 106 lensed galaxies.

We perform the SED fitting in a cell-by-cell base, instead of a pixel-by-pixel base, to obtain the spatial-resolved maps. This is because that the S/N level of individual pixel varies much across the whole images, significant bias in SED modeling result may arise when fitting stellar population models to SED of individual pixel that have marginal or no significance over most or even the entire wavelength range. Therefore, binning the pixel to achieve a sufficiently high S/N level is still a crucial step to obtain feasible spatially-resolved maps, though the bin size can be smaller compared with that from the field survey. To achieve this, we group the pixels into several adjacent cells following the Voronoi two-dimensional binning technique by Cappellari & Copin (2003). The cell size is determined by the criteria that the total S/N within the cell reaches a minimum level of 5 in the synthetic detection images. Then we compute magnification corrected photometry for all the cells in each object (the magnification factor is the median value of the magnification map within the corresponding cell), and fit the spatial-resolved SED by iSEDfit with similar parameter setting mentioned above. Since the strong lensing effect distorts the shape of the PSF, a gaussian approximation to PSF function is not valid for extracting photometry in the source plane. To this end, all the calculations are

performed in the image plane.

3.9 Source Plane Image Reconstruction

The advantages of the lensed images in the image plane are the magnified flux and surface area that provides us insight about the substructure. However, the since the image itself is also highly distorted, such distortion limits our ability to accurately extract their morphological and structural information. Therefore, The study of the morphology and spatial distribution of the sub-galactic structure needs to be performed on the reconstructed images in the source plane.

In simulations, we could obtain lensed images based on the ray tracing techniques used in the giant arc simulation which was briefly discussed in Chapter 2. To simulate a giant arc in the image plane, bundles of light rays are shot from the observer and deflected by the gravitational field in the middle. Those rays which finally hit the background galaxies in the source plane are traced back to where they get deflected in image plane. For a given traced pixel in the image plane, a intensity value which is the bilinear interpolation based on the values of its nearest four pixels in the source plane, is assigned. This value will be multiplied by the magnification factor at that pixel. A lensed image with enhanced flux and surface area is then created in the image plane.

In general, the image reconstruction process is the inverse process of the ray tracing process. However, several challenges make the reconstruction less straight forward. The magnification is not evenly distributed in the image of the lensed galaxies. In most of the case, there exists a dominant direction that receives the largest magnification (in tangential direction). A direct inverse ray-tracing on such unevenly magnified image will result in unevenly distributed spatial resolution in the reconstructed image in the source plane. If we adopt an uniform spatial resolution in the source plane that is the same as that in the image plane, the adjacent pixels with high magnification in the image plane will likely be mapped to overlap with each other in the source plane. To overcome this difficulty, we perform the source plane image reconstruction as follows:

First, we compute an overall magnification factor for each lensed galaxy (the H-band intensity-weighted average magnification) and its square root will be the average linear magnification factor.

Second, we expand the image of the lensed galaxy and the related lensing models by a factor that equals to the average linear magnification factor.

Third, we perform a direct ray-tracing mapping on these magnified image based on the magnified lensing models. A primary delensed image in the source plane is obtained.

Due to the unevenness of the magnification in the lensed image, there will

be some pixels missing in the source plane that correspond to areas of low magnification in the image plane. Therefore a further image interpolation in the source plane is needed. However, a direct image interpolation based on the square grid may not be proper because a square grid in the image plane will be mapped into an irregular grid in the source plane due to the distortion of strong lensing. Instead, we consider to interpolate the image based on a triangular grid in the source plane. The advantage of using triangular grid is that a triangular unit in the image plane remains triangular in the source plane after mapping. The construction of triangular grid is based on the Delaunay triangulation: the intensity value of a given pixel is determined by the values of three closest vertices and the distance to the vertices of the host triangle. As a final step, we convolve the interpolated image by a Gaussian kernel with width = 2 pixel in the source plane to remove the noise and enhance the contrast of the image to the background.

Using the above technique, we reconstruct the spatially-resolved maps for various of physical properties for all 106 lensed galaxies in the source plane. Figure 3.4 lists the representative spatially-resolved maps for stellar mass, star formation rate (SFR) and specific star formation rate (sSFR) of two CLASH highly lensed galaxies in both the image and source planes. The full image catalog can be found in Figure G.1 in Appendix G

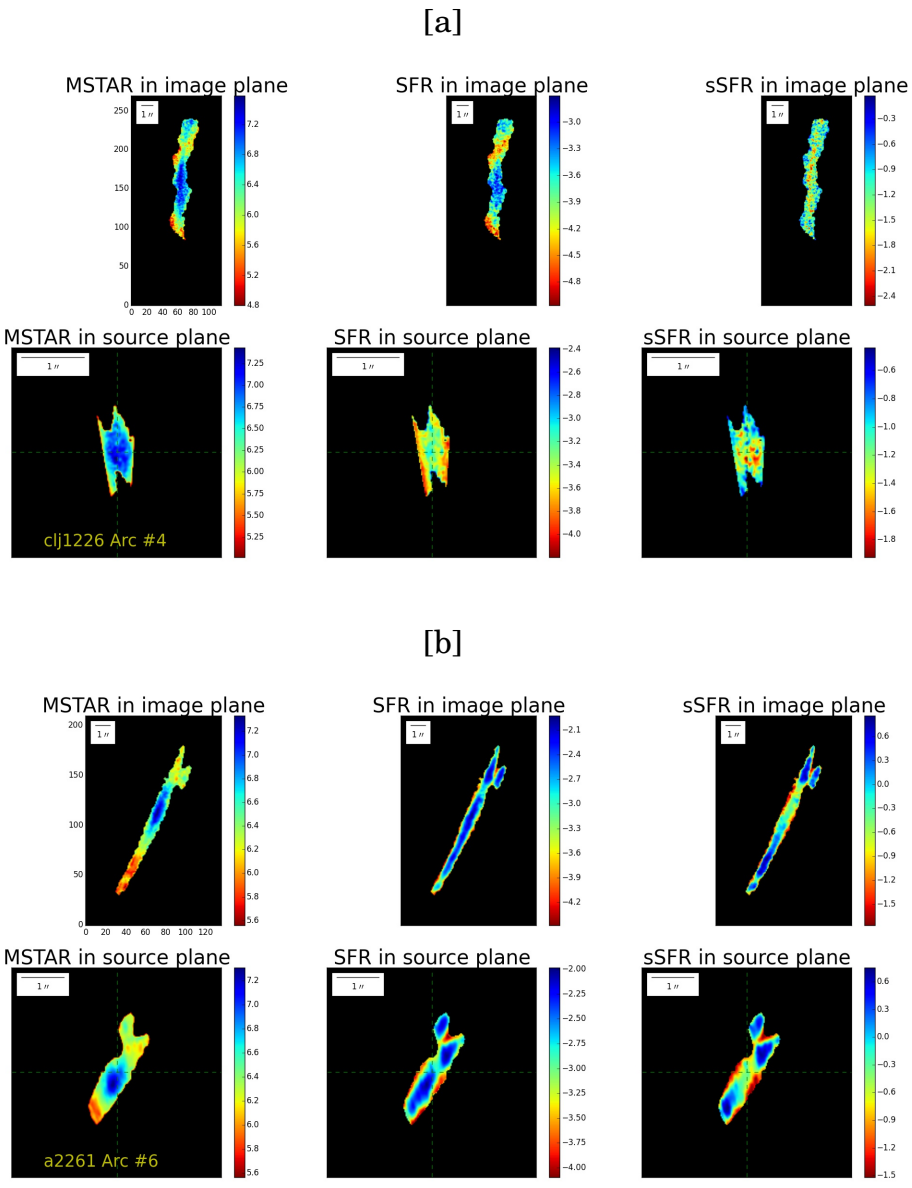


Figure 3.4: The spatially-resolved maps for stellar mass, SFR and sSFR of two representative CLASH lensed galaxies. The upper panels are the maps in the image plane; the bottom panels are the maps in the source plane.

3.10 Arc Morphological Classification

Given the reconstructed images in the source plane, we are able to explore the sub-galactic structure in a higher spatial resolution than that in the image plane. We investigate three key aspects of the sub-galactic structure:

1. the integrated structural properties of the host galaxies, such as their morphology (as determined from objective morphological indicators);

2. the physical properties of sub-galactic clumps with high rest-frame UV flux contrast relative to the smooth galactic background;

3. the radial distribution and radial variation of the physical properties of sub-galactic clumps.

We discuss the morphologies of the lensed galaxies in this Chapter. The physical properties and radial distribution of clumps are discussed in Chapters 4 and 5.

The non-parametric morphological classification parameters such as Gini, M_{20} , asymmetry, and smoothness are widely used to automatically separate galaxies into a few basic categories such as disk, elliptical and irregular. For example, these parameters have been used for galaxy morphology studies with HST and SDSS galaxy surveys (Conselice et al., 2000; Abraham et al., 2003; Lotz et al., 2004; Zamojski et al., 2007; Wang et al., 2012) and for analyses

of high-z galaxies (Abraham et al., 2003; Lotz et al., 2004; Wang et al., 2012). Most of these morphological studies focus on field galaxies and the PSF blurring effect would hinder robust morphological classification of small galaxies and sub-galactic structre at redshifts $z > 2$ because at z$\sim$ 2 the scale of 1 kpc corresponds to an angular scale 0".12, which is close to the 0."08 FWHM of the PSF of the HST F814W filter. As a consequence, a clumpy irregular galaxy might be easily mis-identified as a regular disk galaxy (Rigby et al., 2017). It is not clear how much the morphological indicators depend on the presence of marginally-resolved substructures. On the other hand, strong lensing magnifies the galaxy image and enables the detection and characterization sub-kpc structures at $z > 1$. Therefore, we want to answer the following questions:

Will the enhanced spatial resolution due to the magnification by the strong lensing significantly improve the accuracy of morphological measurements?

Or equivalently,

Are the defined morphological indicators robust to the missing substructures?

To study this, we simulate large amounts of images of lensed galaxies with different magnifications in the image plane and reconstruct them in the source plane by the same procedure as we do to the observational data. We measure the morphological parameters of the reconstructed images in the source plane and assess the uncertainty and bias of these parameters at different magnifi-

cations.

We present a brief review of these morphological indicators in the following sections.

3.10.1 Gini coefficient

The Gini parameter is widely used in the field of economics, where it originated as the Lorenz curve (Lorenz, 1905). It describes the inequality of wealth in a population. If the total flux is uniformly distributed among the pixels, then the Gini value is equal to zero; but if the total flux is unevenly distributed and belongs to only a small number of pixels, then the Gini value approaches a value of unity. We adopt the following definition from Lotz et al. (2004):

$$G = \frac{\sum_i^N (2i - N - 1)|f_i|}{(N-1)\sum_i^N |f_i|}, \tag{3.7}$$

where N is the number of pixels within the images and f_i are the fluxes for each pixel sorted in ascending order with $|f_1| \leq |f_2| \leq ... \leq |f_N|$.

3.10.2 2nd-order Moment of Light Distribution M_{20}

Lotz et al. (2004) define the total second-order moment M_{tot} as the flux

in each pixel multiplied by the squared distance to the centre of the source, summed over all the selected pixels:

$$M_{tot} = \sum_{j}^{N} M_j = \sum_{j}^{N} f_j \cdot ((x_j - x_c)^2 + (y_j - y_c)^2), \tag{3.8}$$

where x_c and y_c are the galaxy's center and f_i are the fluxes for each pixel sorted in descending order with $|f_1| \geq |f_2| \geq ... \geq |f_N|$.

M_{20} is defined as the normalized second-order moment of the relative contribution of the brightest 20% of the pixels. The higher M_{20} is, the more irregular morphology a galaxy would have. To compute M_{20}, we rank-order the image pixels by flux, calculate f_i over the brightest pixels until their sum equals 20% of the total flux, and then normalize by M_{tot}:

$$M_{20} = \log 10 \left(\frac{\sum_j M_j}{M_{tot}} \right) \text{while} \sum_j f_j < 0.2 \sum_{j}^{N} |f_j| \tag{3.9}$$

3.10.3 Asymmetry and Smoothness

The asymmetry parameter, A, quantifies the degree to which the light of a galaxy is rotationally symmetric. The parameter A is measured by subtracting the galaxy image rotated by 180 from the original image (Abraham et al., 1996;

Conselice et al., 2000):

$$A = \frac{\Sigma_{i,j}|I(i,j) - I_{180}(i,j)|}{\Sigma_{i,j}|I(i,j)|} \tag{3.10}$$

where I is the galaxy's original image and I_{180} is the image rotated by 180° about the galaxy's central pixel.

The smoothness parameter can be used to identify any clumpy flux distribution. By smoothing the galaxy image with a filter of width σ, high frequency structures can be removed from the image. At this point the original image is subtracted from this newly smoothed, lower resolution image and a residual map that has only high frequency components of the galaxy's flux distribution is created. The flux of this residual image is then summed and divided by the total flux of the original galaxy image in order to find its smoothness value:

$$S = \frac{\Sigma_{i,j}|I(i,j) - I_S(i,j)|}{\Sigma_{i,j}|I(i,j)|} \tag{3.11}$$

where I_S is the galaxy's image smoothed by a boxcar of width 0.25 r_p and r_p is the Petrosian radius.

In sum, different morphological parameters are sensitive to different morphological type:

Gini coefficient is most sensitive to the central flux concentration (central bulge, star clusters, etc) or clumpy structure;

M_{20} is most sensitive to irregular flux distribution (tidal feature, merger remnant, asymmetric spiral arms, etc);

Asymmetry is most sensitive to the asymmetric flux distribution (tidal feature, merger remnant, asymmetric spiral arms, etc);

Smoothness is most sensitive to the clumpy structure.

3.10.4 Correcting for Bias in the Morphological Classification

Aside from the normal source of uncertainty when measuring the morphological parameters in the field survey, such as PSF blurring, pixel binning, sky noise, etc, extra uncertainty such as any inaccuracy in the lens models is introduced in the process of image reconstruction in the source plane. To quantify the overall uncertainty, we have simulated sets of galaxy images by taking these effects taken into account. We first select 30 galaxy images with various morphologies from realistic Illustris simulation (Vogelsberger et al., 2014), in which important gas physics such as feedback heating and stellar winds has been properly implemented. The Illustris galaxy images have pixel resolution ~ 0.006" per pixel and we adopt these galaxies as our source galaxy images. We then respectively bin each image down to resolutions of 0.065", 0.033", 0.022" and 0.016" per pixel, which approximately correspond to $\times 1$, $\times 2$, $\times 3$ and $\times 4$

resolution of CLASH images (Since the maximum overall magnification factor of CLASH giant arc is $\sim$ 20, we only consider the linear magnification factor up to 4). We place these binned image around the caustic lines in the source plane that correspond to the magnification factors = 2,4,9 and 16 for 20 X-ray selected CLASH clusters at $z = 2$ [1]. Using the ray-tracing technique, we create the images of the lensed arc in the image plane at each magnification (resolution). We then modify the images by adding the observational effects: first, convolve the image with PSF function for HST F160W filter; then bin the image down to the CLASH pixel scale and add Gaussian noise that matches the real CLASH images, to create the simulated "observed" images in the image plane. We perform the same image reconstruction procedure on these simulated images as we do on the real CLASH images. All the morphological parameter measurements are performed on the reconstructed images in the source plane. In addition, we create 5 simulated realizations for each source galaxy (out of 30), magnification factor (out of 4) and CLASH cluster (out of 20). We have obtained $30 \times 20 \times 5 \times 4 = 12000$ simulated galaxy images for the subsequent analysis. We calculate the ratio of the simulated "observed" morphological parameters to the "real" value as a function of magnification factor for all the source galaxies. For each magnification factor, we calculate the mode value and the 1σ error of the distribution of the fractional bias. The results are shown in

[1]The lensing efficiency with placing all the background galaxies at $z = 2$ is similar to that with a realistic redshift distribution. See Chapter 2

Figure 3.5. For magnification factor > 4 (typical CLASH lower limit), the average fractional bias for Gini coefficient, M_{20}, asymmetry and smoothness are approximately: $\frac{\Delta G}{G} = 0.05$, $\frac{\Delta M_{20}}{M_{20}} = 0.2$, $\frac{\Delta A}{A} = -0.2$ and $\frac{\Delta S}{S} = -0.4$, respectively. For Gini coefficient and smoothness, as expected, the fractional bias and its error bar decreases as the magnification factor increases, and the decrement is modest (a factor of 2) when the magnification increases from 1 to 16. We do not observe any trends for M_{20} and asymmetry toward the higher magnification, though the scatter for M_{20} decreases for a factor of 3 as it reaches the highest magnification. The asymmetry appears to have larger bias at higher magnification. This is likely due to that the clumpy structure is also asymmetric, which will increase the degree of asymmetry at higher resolution whereas at lower magnification, the galaxy become more symmetric with clumpy structure smoothed out. The observed Gini coefficient is biasing high, whereas the measurements of the other three parameters are biasing low (M_{20} is always negative so the fractional bias is positive). This is not surprising because both PSF convolution and sky noise tend to dim the central flux and to lower the central average surface brightness. Thus both the measured Petrosian radius and Gini coefficient are larger. Moreover, both effects tend to smooth or erase the clumpy substructures and stand out the brighter structures, therefore to lower the asymmetry and smoothness. Such effects on Gini coefficient and M_{20} appears to depend on the specific morphology: Gini coefficient (M_{20}) would in-

crease (decrease) if the bright structures are close to the central region and vice versa. It also explains the large scatter at lower magnification for Gini coefficient and M_{20}.

Given the fractional bias and its scatter, we conclude that the improvement on the accuracy of the morphological parameter measurements is limited with magnification by strong lensing. In another words, the standard morphological indicators are robust to the presence smaller sub-galactic scale, marginally-resolved structures. Our results are consistent with Rigby et al. (2017), who also find the morphological parameters are not sensitive to the magnification.

3.11 Results of Arc Morphological Classification

3.11.1 Morphological Distribution of CLASH Sample and Implications

The Gini-M_{20} plane has been proved as a efficient tool to classify galaxies with normal morphology (late and early types) and highly irregular, disturbed morphology (merger remnant) (Lotz et al., 2004, 2008). Therefore, it is essential to investigate where the CLASH lensed galaxies lie in this Gini-M_{20} plane

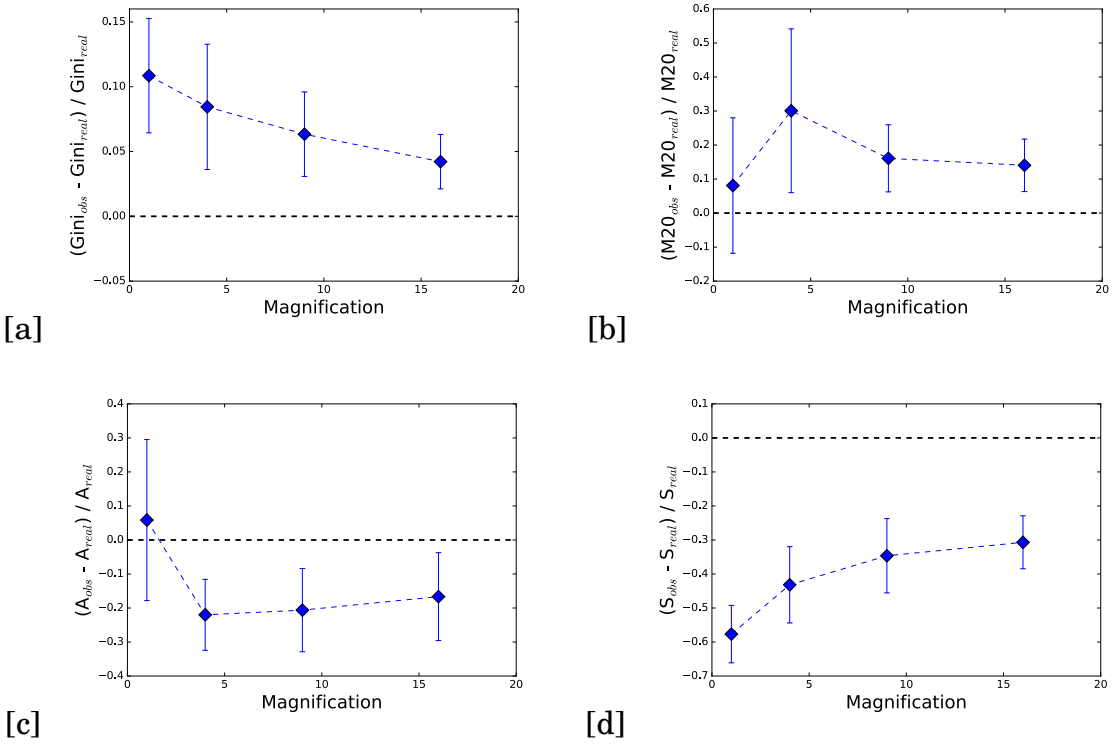


Figure 3.5: Figure (a), (b), (c) and (d) show the average fractional bias of Gini coefficient, M_{20}, asymmetry and smoothness parameters as a function of the increasing magnification factor. The error bar denotes 1σ error.

and to infer their morphology. To be consistent with the previous studies (Lotz et al., 2004), we select a subsample with rest-frame optical images available. We perform image reconstruction of the selected rest-frame 6000 $\mathring{A}$ images and measure their Gini coefficient and M_{20} in the source plane. Figure 3.6 shows the distribution of the CLASH lensed galaxies in the Gini-M_{20} after correcting the magnification bias. The blue and green lines separate the parameter space into three regions: late type, early type and irregular/merger galaxy (Lotz et al., 2008). Most of the galaxies in this subsample fall in the irregular region and their positions are consistent with the local ULIRGs and high-z LBGs. There are only three galaxies in the subsample that are in or close to the late and early type regions. As a double check, we also examine their positions on the color-mass plane and SFR-mass plane (Figure 3.7). Most of the CLASH galaxies are in the blue cloud region and the only galaxy falls into the red sequence is the one which lies in the early type region in Gini-M20 plot. Most of the CLASH galaxies follow the star-forming main sequence very tightly and the only red galaxy in CLASH subsample falls significantly below the star-forming main sequence as expected. The morphology of CLASH subsample is consistent with that the galaxies at $z \sim 1$ typically have irregular morphology and lack of a prominent bulge.

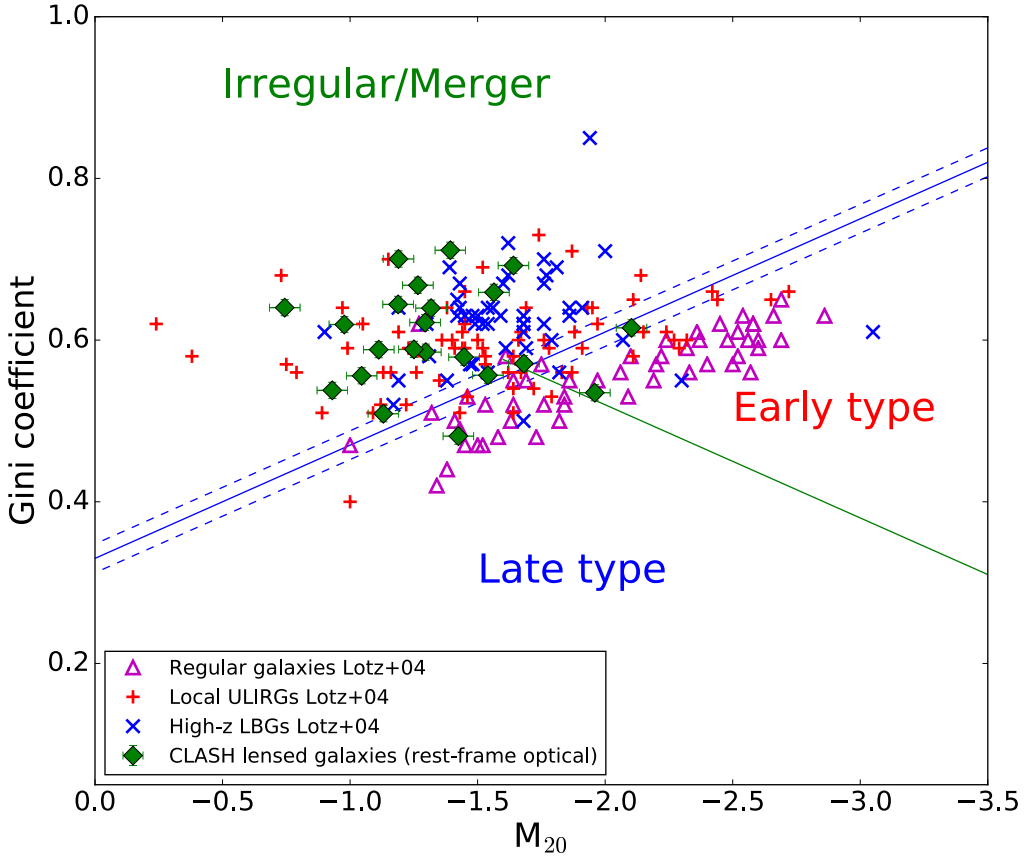


Figure 3.6: CLASH arc subsample with rest-frame optical image available in Gini-M_{20} plot. Blue solid line is the boundary of galaxy with regular and irregular morphology and the blue dashed lines are 1σ error of the boundary; the green solid line is the boundary of late and early type galaxy. The error bar denotes 1σ error. Three galaxy samples are shown for comparison: the cyan triangles are the local regular galaxies; the red "+" are the local ULIRGs and the blue cross are the high-z LBGs. Samples are taken from Lotz et al. (2004).

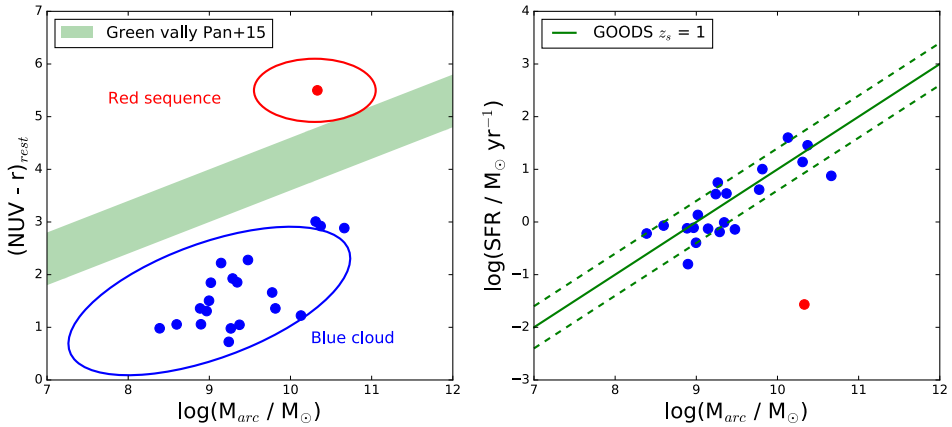


Figure 3.7: Left panel shows the CLASH arc subsample with rest-frame optical image available in the (NUV - r) VS stellar mass plane. The red and blue ellipses denotes the regions of red sequence and blue cloud. The green shaded region marks the position of green vally from Pan et al. (2015). The right panel lists the CLASH subsample in SFR VS stellar mass plane. The green solid line denotes the star-forming main-sequence at $z \sim 1$ and the green dashed lines are 1σ scatter. The blue points are CLASH galaxies with (NUV - r) < 4 and red points are CLASH galaxies with (NUV - r) $>= 4$.

3.11.2 Morphology Comparison Between Rest-frame UV and Optical Images

It is well known that the irregularity of the galaxy morphology at low redshift depends on the specific filter: the galaxy morphology observed in bluer filter tends to be more irregular than that in the redder filter (Lotz et al., 2004). Such "morphological bias" is expected to be less significant if the galaxies at high redshift are intrinsically unsettled and show irregularity in rest-frame optical images (Elmegreen et al., 2009). We divide the rest-frame optical CLASH sample into subsamples at two redshift ranges: $0 < z_s < 1.5$ and

$1.5 < z_s < 2.5$, and compare the distribution of the rest-frame UV and stellar mass weighted with that of rest-frame optical on the morphological parameters to assess whether their morphological difference reduces as redshift increases. Figure 3.8, 3.9 and 3.10 show the comparisons. For the full rest-frame optical sample and two subsamples, we perform two KS tests on Gini and M_{20}: one between the rest-frame UV and rest-frame optical distributions; the other between the stellar mass weighted and the rest-frame optical distributions. For the full sample (Figure 3.8), the distribution of rest-frame UV Gini is similar to that of rest-frame optical and a p-value of 0.71 indicates that we can not reject the hypothesis that they are drawn from the same parent distribution. We can marginally reject the null hypothesis that the distribution of rest-frame UV and rest-frame optical M_{20} are drawn from the same parent distribution with a p-value of 0.013. The comparison at $0 < z_s < 1.5$ (Figure 3.9) is similar and only the rest-frame UV and rest-frame optical Gini show similarity in their distributions. At $1.5 < z_s < 2.5$ (Figure 3.10), the distribution of rest-frame UV and rest-frame optical on both Gini and M_{20} are very similar, and p-values of 0.9969 indicate that we can not reject the hypothesis that they have the same distribution at 5σ level, though such high p-values are also likely due to the small sample size. We can marginally reject the hypothesis that the stellar mass weighted distribution is same with the rest-frame optical distribution in all the case, but on Gini at $1.5 < z_s < 2.5$ with a p-value of 0.534. These results

give us hints that: first, the Gini coefficient is less affected by the morphological bias; second, the morphological difference between the rest-frame UV and optical images appears to decrease as the redshift increases. Therefore, we compute the rest-frame UV Gini coefficient and M_{20} of the lensed galaxies at $z_s > 2.5$ as proxies of their rest-frame optical morphological indicators, and combine them with the rest-frame optical morphological measurements at $z_s < 2.5$. After correcting for the magnification bias, we construct a full lensing based morphological catalog spanning the redshift from 0.5 to 6.2. We will present the results of morphological evolution based on this catalog in next section.

3.11.3 Morphology Distribution at Different Redshift

Provided the full morphological catalog over broad redshift range, we now compare the morphology of CLASH lensed galaxies at different redshifts. Figure 3.11 shows the morphology distributions at 5 redshifts color-coded by their stellar mass. At all redshifts, most of the CLASH galaxies lie in the irregular region. At $z_s \sim 1$, the scatter in Gini coefficient and M_{20} is $\sim$ 0.4 and 1 dex. The scatter in Gini does not change much at higher redshifts; whereas the scatter in M_{20} increase to $\sim$ 2 - 2.5 dex, which indicates the increasing of

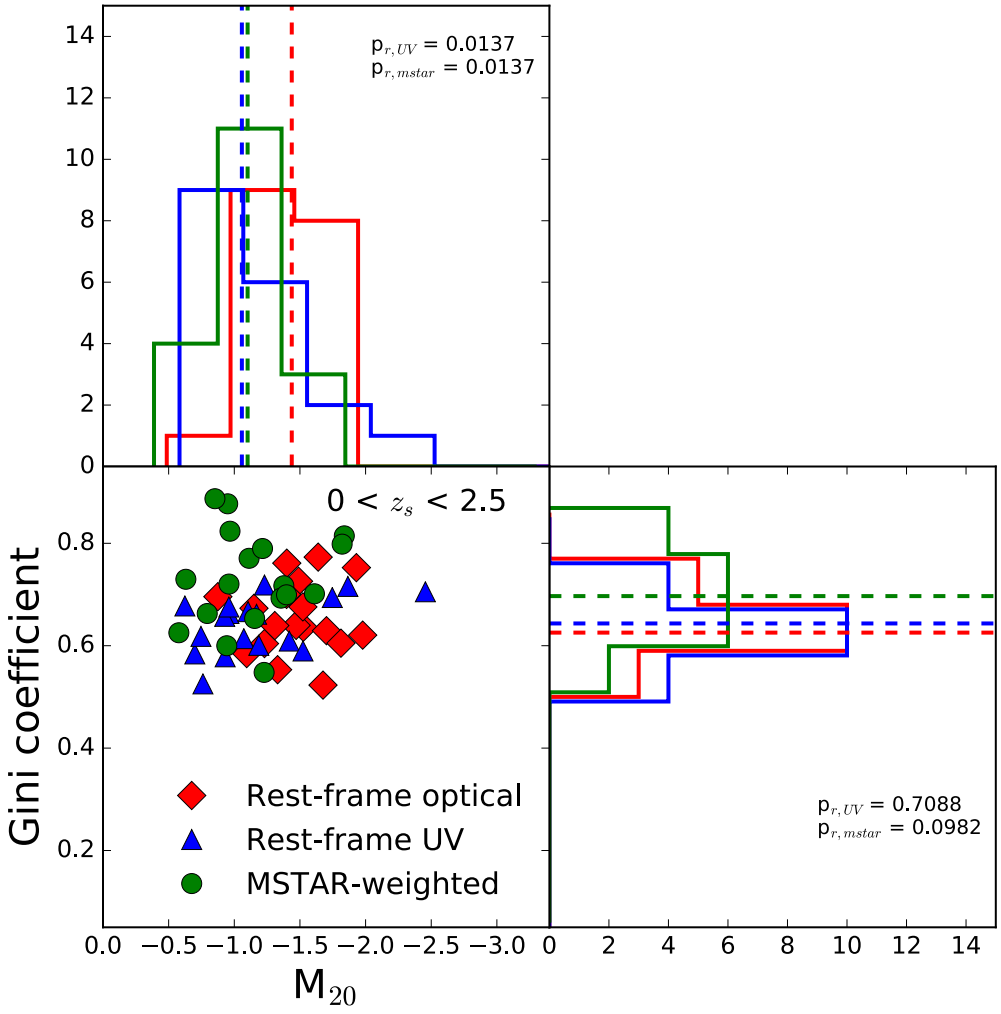


Figure 3.8: Comparison of the rest-frame UV, optical and stellar mass weighted morphology at $0 < z_s < 2.5$. The histograms are marginal distribution of Gini coefficient and M_{20}. The dashed lines with different colors denote the median value of three distributions.

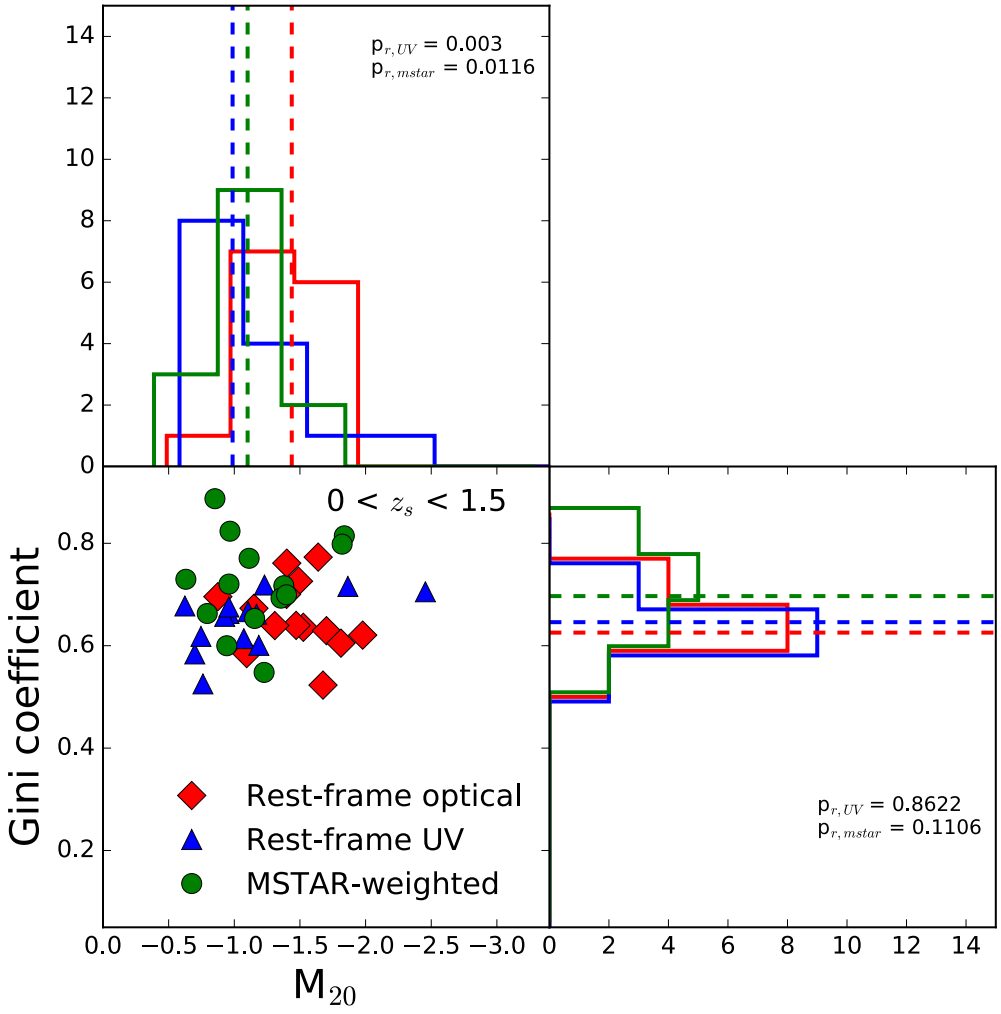


Figure 3.9: Comparison of the rest-frame UV, optical and stellar mass weighted morphology at $0 < z_s < 1.5$. The histograms are marginal distribution of Gini coefficient and M_{20}. The dashed lines with different colors denote the median value of three distributions.

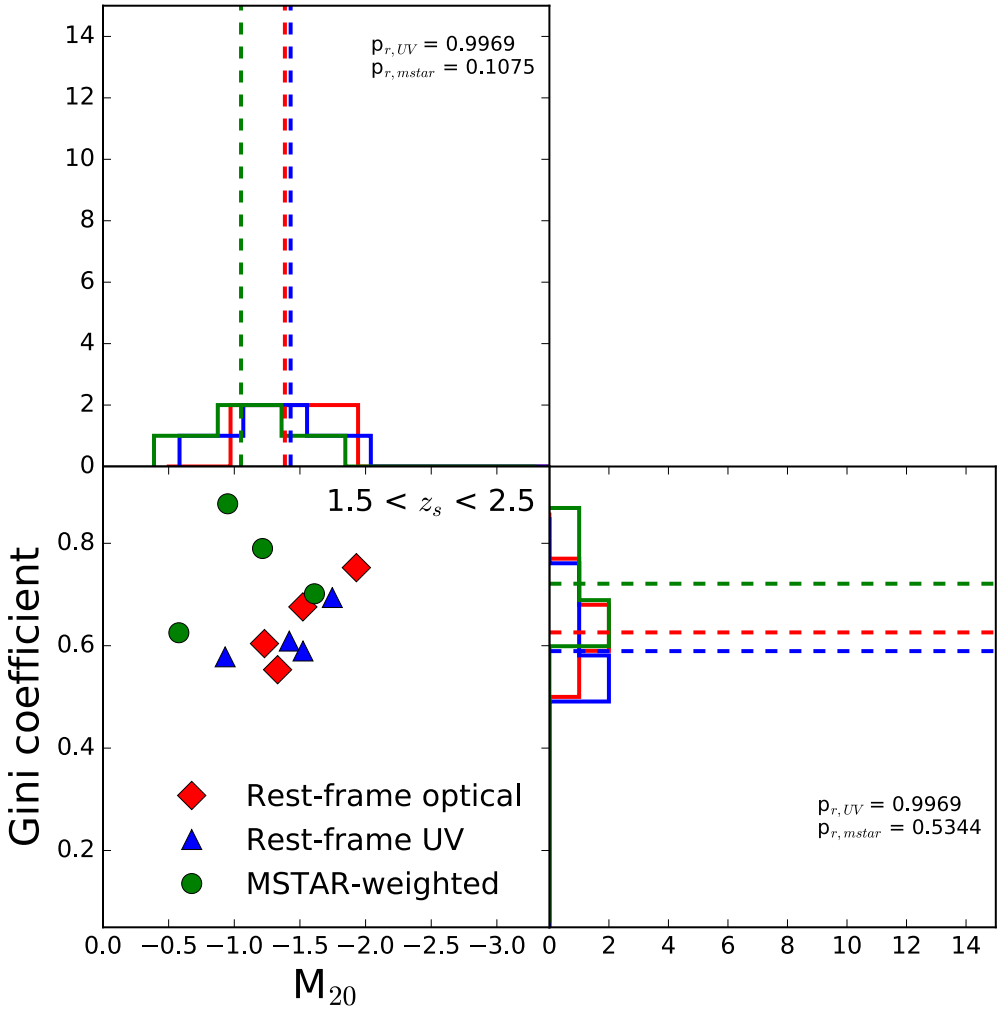


Figure 3.10: Comparison of the rest-frame UV, optical and stellar mass weighted morphology at $1.5 < z_s < 2.5$. The histograms are marginal distribution of Gini coefficient and M_{20}. The dashed lines with different colors denote the median value of three distributions.

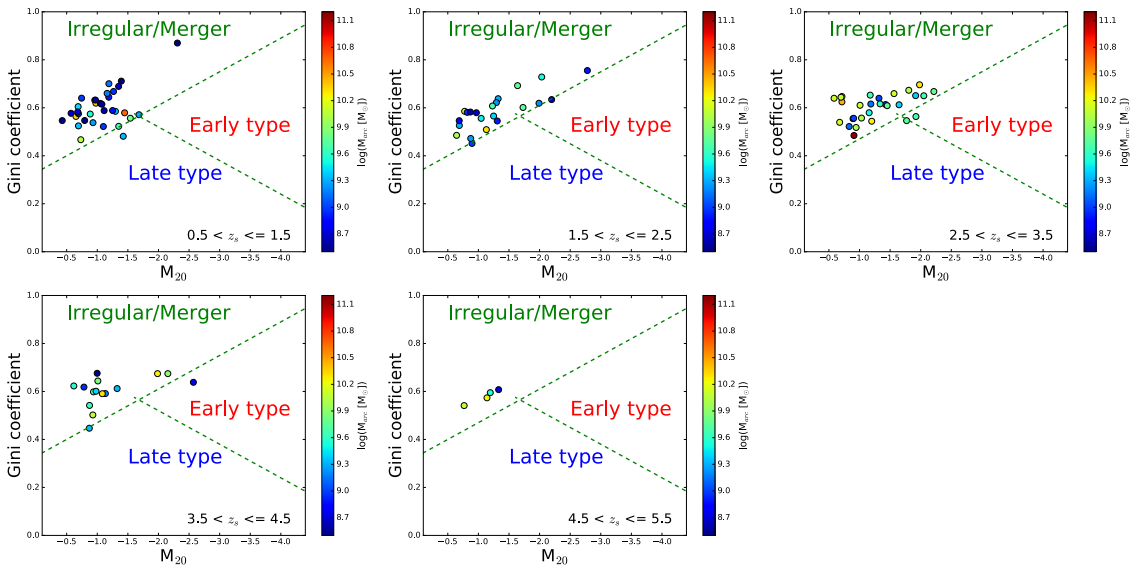


Figure 3.11: Morphology comparison in Gini-M_{20} plane for 5 redshift bins. The dashed lines separate the plane into different morphological types (Lotz et al., 2008). The color bar denotes their stellar mass.

the diversity of galaxies with different morphology. In addition, there are more galaxies at $1.5 < z_s < 4.5$ that have lower M_{20} (more regular morphology) than at $z_s \sim 1$, which indicates that it is more difficult to select galaxies with organized morphology at lower redshift via lensing. This is likely due to that the galaxies with regular morphology at lower redshift are typically more massive (less abundant) than those have been selected by lensing.

We further explore whether the morphology of lensed galaxies depends on their mass at high redshift. We divide the full sample into two subsamples by the overall median stellar mass $\mathrm{M}_\star^{\mathrm{med}} = 10^{9.5}\mathrm{M}_\odot$: the low mass and high mass sample and compute the Gini coefficient and M_{20}, respectively. Figure 3.12 and 3.13 show the distributions of Gini coefficient and M_{20} for three samples. Due

to lack of less massive galaxies at $z_s \sim 5$, we only plot the full and high mass sample at $z_s \sim 5$. We do not observe any significant difference in measured morphological parameters between the high and low mass sample, nor do we detect any noticeable evolution in the Gini coefficient and M_{20}. Therefore, we conclude that the morphology of CLASH lensed galaxies at high redshift is less sensitive to their host galaxy mass. It appears in the face value in stark contrast with the case at lower redshift that more massive galaxies usually have more organized morphology. However, these nearby massive galaxies typically have stellar mass $10^{10} - 10^{12} M_{\odot}$ whereas most of CLASH lensed galaxies have stellar mass less than $10^{10} M_{\odot}$. Therefore it is likely that the morphology of galaxies would be significantly affected by their potential well only after their assembled mass reaches above some minimum value.

We plot the morphological comparison for full, high and low mass sample at different redshift in Figure 3.14 for reference.

3.12 Summary and Conclusions

In this chapter, we study the integrated physical properties and morphology of strongly lensed galaxies at high-z. The study is based on a sample which is composed of 106 lensed galaxies which have been detected by our arc-finder from 20 X-ray selected CLASH clusters. We measure their photometric red-

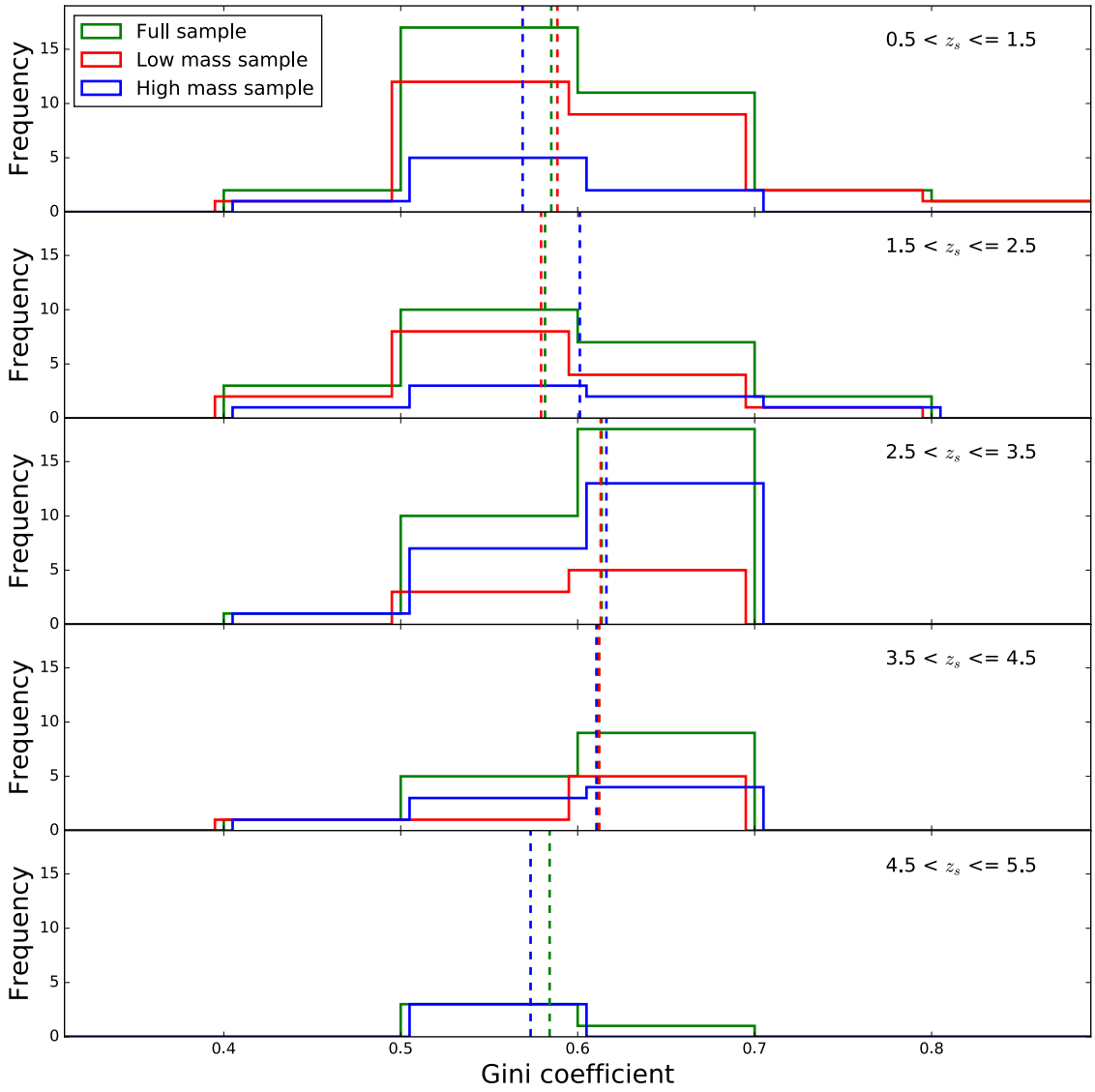


Figure 3.12: The distributions of the Gini coefficient of three samples for 5 redshift bins. The colored dashed lines are the median values of the three distributions.

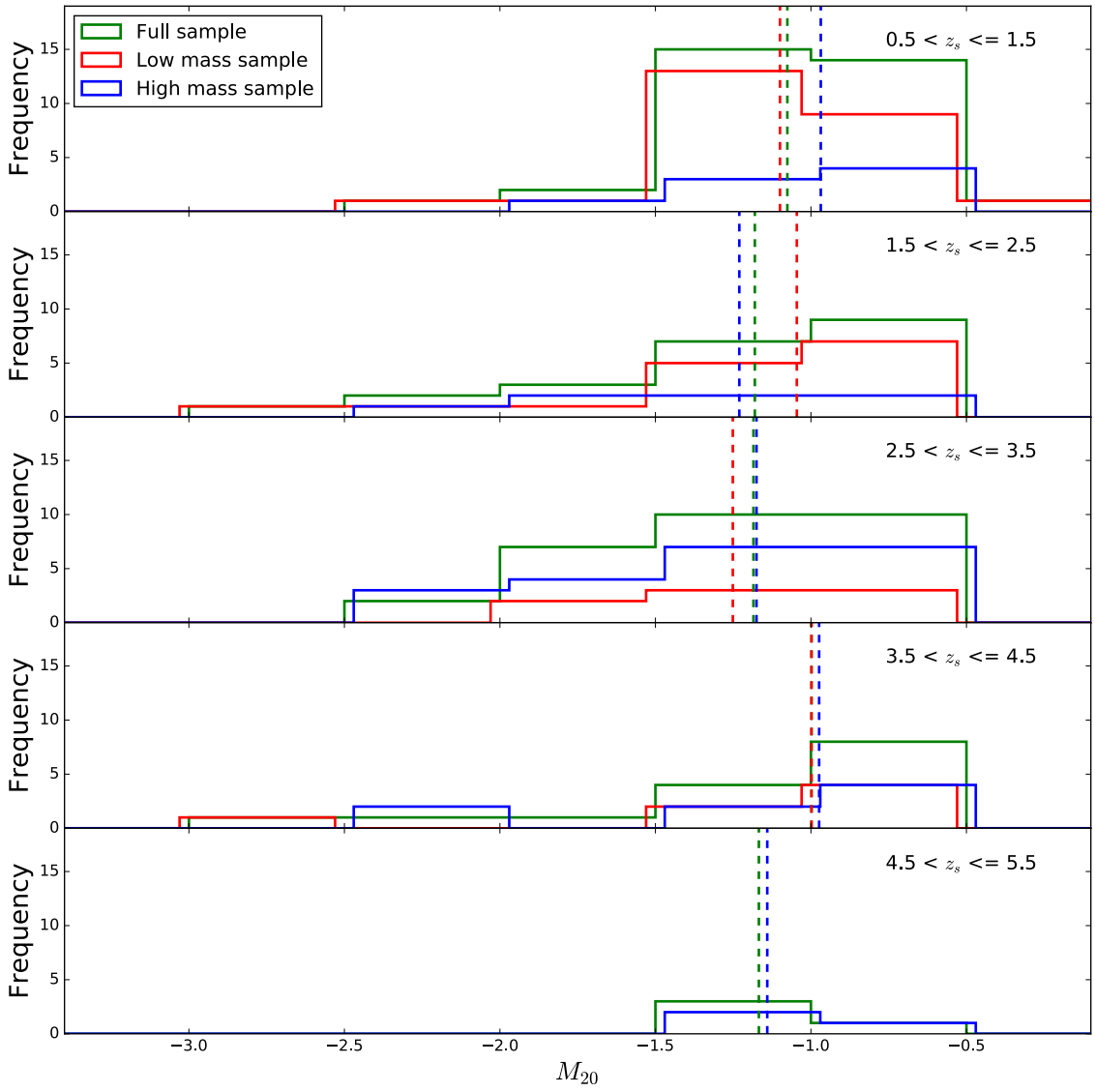


Figure 3.13: The distributions of M_{20} of three samples for 5 redshift bins. The colored dashed lines are the median values of the three distributions.

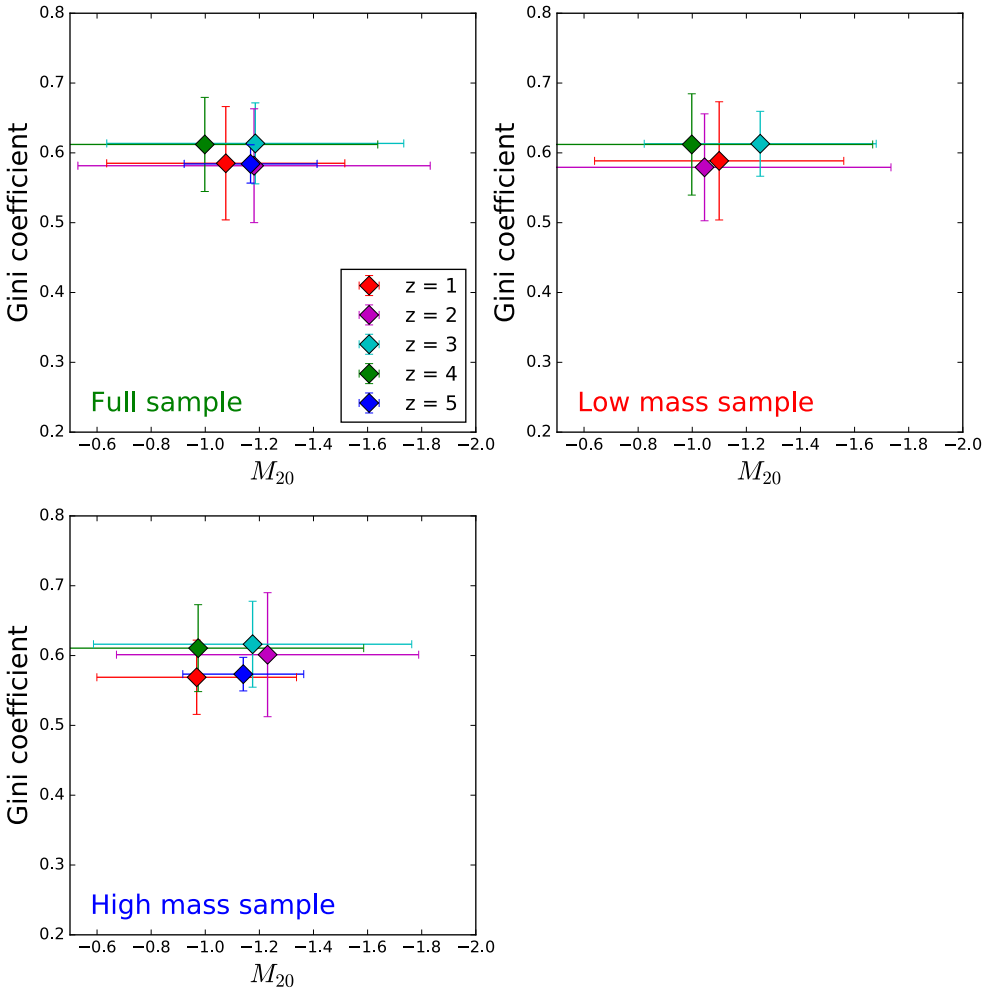


Figure 3.14: The median value of Gini coefficient and M_{20} at 5 redshifts for three samples. The error bar denotes the 1σ standard deviation.

shift and extract the photometry after correcting for lensing magnification. We use iSEDfit to model the SEDs of the lensed galaxies to infer their integrated physical properties such as stellar mass, SFR and sSFR. We compare the distributions of stellar mass and SFR of our sample with the CANDELS UDS survey. We find that our sample preferentially selects galaxies with intermediate mass and midl SFR, with a median stellar mass of $10^{9.4}\mathrm{M}_{\odot}$ and SFR of $4\mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$. We compute the sampling volume as a function of redshift and lensing magnification. We perform the cell-by-cell based photometry extraction and SED fitting, and create the spatially-resolved maps for various physical properties. We study the morphology of the lensed galaxies by exploring their position on Gini-M20 plane. We find most of the CLASH lensed galaxies lie in the irregular region. We find that the morphological difference of galaxies between the rest-frame UV and optical images at $z_s > 1.5$ is smaller than that at $z_s < 1.5$. We construct a full morphological catalog by combining the rest-frame optical morphological parameters at low-z and rest-frame UV morphological parameters at high-z. We compare the parameters at different redshifts and find that at high-z, galaxies generally have irregular and clumpy structure and no significant evolution in morphology has been detected. We find that for galaxies with mass less than $10^{10}M_{\odot}$ at high redshift, the morphology is not much affected by their mass. We suggest that there could be some critical mass above which the potential well of the galaxies starts to significantly affect their morphology.

Chapter 4

Galaxy Evolution on the Sub-Galactic Scale at $1 < z < 5$

4.1 Introduction

Deep field survey with HST have revealed that majority of galaxies at $1 < z < 3$, the peak of the star forming activity (Madau & Dickinson, 2014), exhibit irregular and clumpy morphology in the rest-frame UV images (Elmegreen et al., 2007; Guo et al., 2012, 2015; Shibuya et al., 2016), rest-frame optical images (e.g., Elmegreen et al. (2009); Forster Schreiber et al. (2011)), rest-frame optical line emission images from NIR integral field spectroscopy (Genzel et al., 2008, 2011) or CO line emission images of high-z lensed galaxies (e.g., Jones et al. (2010); Swinbank et al. (2010)). Such ubiquitous irregular, clumpy mor-

phology at high redshift is in stark contrast with the more organized galaxy morphology in the local Universe. To reveal how this transition occurs is crucial to complete our understanding on galaxy evolution. Therefore, a detailed study of the clumps, as the characteristic feature of the clumpy galaxies at high redshift, is the key step in expanding our understanding of the role that these sub-galactic components play in forming and establishing the galaxy structure at current epoch.

The typical stellar mass ($M_\star$) of high-z clumps is $10^7-10^9 M_\odot$ (e.g., Elmegreen et al. (2007); Guo et al. (2012)), and the typical size is $\sim$ 1 kpc or less (e.g., Elmegreen et al. (2007); Forster Schreiber et al. (2011); Livermore et al. (2012, 2015); Rigby et al. (2017); Vanzella et al. (2017)), which are much more massive and larger than the local Giant Molecular Clouds (GMCs), which typically have mass $\sim 10^5 - 10^6 M_\odot$ and size $\sim$100 pc (Larson, 1981; Roman-Duval et al., 2010). The formation mechanism of high-z massive clumps is still unclear. In a widely held hypothetical senario based on theoretical works and numerical simulations, the high surface density of the cold gas Σ_g makes the Toomre parameter of the gas disc $Q \sim \sigma\Omega/(\pi G\Sigma_g) \leq 1$ (Toomre, 1964), where σ is 1D gas velocity dispersion and Ω is the angular frequency; a Q parameter that is lower than unity will make the gas disc unstable and thus the massive clumps could form through such violent disk instability (VDI) in gas-rich turbulent disks (e.g. Noguchi (1999); Immeli et al. (2004a,b); Elmegreen et al. (2009);

Dekel et al. (2009b); Ceverino et al. (2010, 2012)). This senario is supported by the observations which show that high-redshift galaxies are gas-rich, with a gas-to-baryonic fraction of 20%-80% (e.g., Erb et al. (2006); Genzel et al. (2008); Tacconi et al. (2008, 2013); Forster Schreiber et al. (2009); Daddi et al. (2010)), and also by the observations of gas kinematic (Genzel et al., 2008, 2011), which show that many clumpy galaxies have underlying rotating disks. This scenario has intriguing implications on galaxy evolution: if the clumps can survive the stellar feedback to be long-lived, they could be driven towards the galactocenter by the interactions between the clumps, dynamical friction against the underlying disks as well as the gravitational torque exerted by the underlying disk in a few dynamical timescale, and eventually coalesce into a bulge (e.g., Bournaud & Elmegreen (2009); Dekel et al. (2009b); Murray et al. (2010); Genel et al. (2012))

In classic picture of galaxy formation, the bulge built-up is accomplished via major mergers of galaxies, which could efficiently drive gas to the galactocenter and produce a starburst to contributes to the central concentration (Toomre & Toomre, 1972; Milos & Hernquist, 1996; Hopkins et al., 2012). However, to explained the abundance of the local galactic bulges in the frame of galaxy merger, a much higher frequency of galaxy merger are required than that has been observed (Lotz et al., 2011). Meanwhile, VDI and clump migration based bulge build-up mechanisms have the advantage of not requiring frequent

galaxy merging at high redshift. However, it is still unclear whether the massive clumps can survive from the stellar feedback in the process of migration. Different simulations give different predictions on the lifetime of clumps, from 50 Myr to 500 Myr (Murray et al., 2010; Genel et al., 2012; Bournaud et al., 2014; Oklopcic et al., 2017; Mandelker et al., 2014, 2017). Therefore, accurate measurements on physical properties of clumps are vital to set up constraints on the theoretical models such as clump migration and galactic bulge mass assembly.

In this study, we extend the previous pioneering works on the high-z clumps (Elmegreen et al., 2007; Jones et al., 2010; Swinbank et al., 2010; Forster Schreiber et al., 2011; Guo et al., 2012; Livermore et al., 2012, 2015; Wuyts et al., 2014; Rigby et al., 2017; Vanzella et al., 2017), based on a larger sample of 193 detected clumps spanning the redshift range $0.5 < z_s < 6$. The high spatial resolution achieved by strong lensing combined with the availability of multi-wavelength image data enable us to accurately measure the physical properties of clumps such as photometric redshift, stellar mass, star formation rate (SFR), specific star formation rate (sSFR), color, age, metallicity as well as their radial distributions. We use these measurements to investigate the formation mechanism of the clumps, its connection with host galaxies and other stellar systems, the feasibility of clump migration based bulge build-up mechanism.

4.2 Clump Detection and Characterization

4.2.1 The Input Giant Arc Sample

We detect clumps from the same sample of CLASH lensed galaxies discussed in Section 4.2.1. In some cases, a giant arc with high magnification and large l/w ratio may itself contain multiple images of the same background galaxy. This is mainly due to the merging of smaller arclets with opposite parity (mirror symmetry) that lie on different sides of the critical line in the image plane. Thus a simple sum of all detected arcs could result in some duplicate counting. Because a common characteristic of these cases is that the critical lines will intercept the main body of the giant arcs, we attempt to correct for this issue as follows:

First, we plot the segmentation maps (see Chapter 2 and Appendix A-F) of all the giant arcs along with the critical lines at their measured photometric redshift.

Second, we visually select the giant arcs in which interception of critical line occurs, and divide the father giant arc as two arclets based on the position of the interception.

Third, we re-calculate the overall H-band flux-weighted magnification fac-

tor for each arclet and select the one with larger magnification as the new representative of the father giant arc.

Figure 4.1 shows examples of giant arcs intersected by critical lines.

We reconstruct these newly selected arclets in the source plane, and obtain a new sample of 106 reconstructed images in the source plane for the subsequent studies.

4.2.2 Clump Definition and Detection Criteria

Given the sample of reconstructed images, we devise an automated algorithm to identify the clumps in the source plane. Using automated clump-finder has several advantages such as: less subjective, results reproducible and the availability to objectively measure the clump completeness, which we will discuss in the next section. Although there have been several studies on clumps in both low and high redshift galaxies, a consensus on the definition and detection criteria for clump has yet to come into being. A popular choice is to select clump based on the ratio of its rest-frame UV flux to the overall UV flux of the host galaxy. One can set a stringent criteria to select brighter clumps (e.g. $> 8\%$ Guo et al. (2015)), or include the fainter ones (e.g. $> 1\%$ Boada et al. (2015)). One can also set alternative criteria on the local flux contrast (Soto et al., 2017). Overall, the clumps are defined as individual sub-galactic star-forming regions with intense star formation which is dictated by their rest-frame UV luminos-

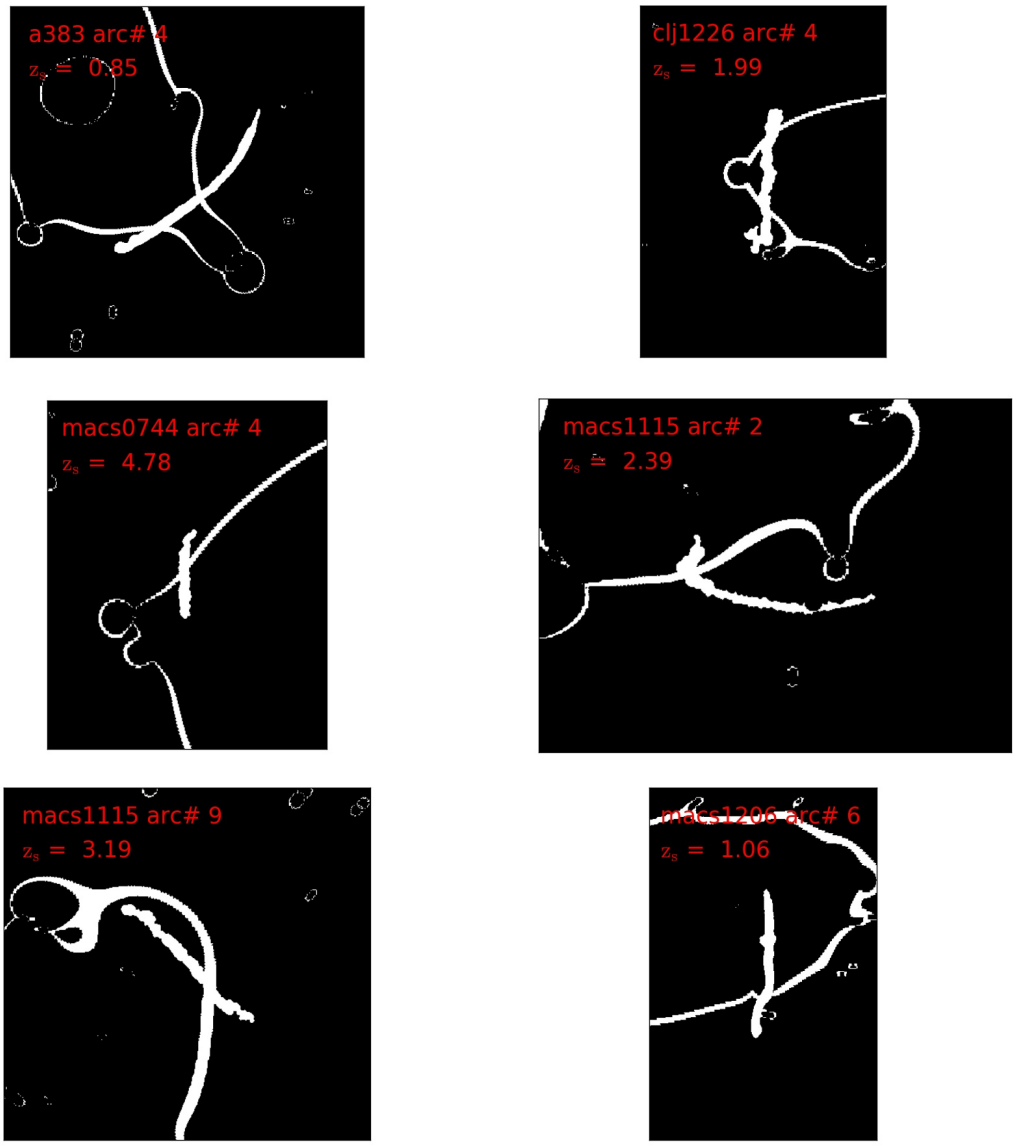


Figure 4.1: The segmentation map of the giant arcs with the intercepting critical lines at their measured photometric redshifts.

ity. In this study, we characterize and detect clumps as follows:

(i) We choose to detect clumps in their rest-frame $2200\mathring{A}$ other than in rest-frame filters at shorter wavelength that better trace the formation of young stars (e.g. 1500 $\mathring{A}$) because the WFC3/UVIS filters in CLASH survey do not guarantee a high S/N for clump detection in rest-frame $1500\mathring{A}$ at low redshift. The rest-frame 2200 $\mathring{A}$ filter includes the starlight from older stellar populations that allows us to probe the internal color and sSFR variance of the clumps, which is crucial for the assessment of the lifetime of clumps.

(ii) All the clump detection have been performed in the reconstructed images in the source plane. The detection of clumps is composed of three main steps: estimation of background and fluctuation; primary segmentation by S/N cut and final segmentation by area cut.

(iii) We estimate two sets of background levels and related fluctuations. We first measure the sky background level and noise outside the clump aperture. We iteratively clipped out the brightest pixels that fall out of 3σ level to the median value, until the value of the remaining pixels converges within 3σ to its median value. We adopt the median value I_s and the standard deviation σ_s as the sky background level and noise. We then repeat the above process within the clump aperture to estimate the local background level I_g and its fluctuation σ_g.

(iv) We first mask out all the pixels below the intensity threshold

$$I_{thr} = I_g + 4\sigma_g. \tag{4.1}$$

to obtain primary clump segmentation. Comparing with the clump detection in field galaxies, the possible inaccuracy in modeling the lens mass distribution would bring in larger uncertainty in the spatial distribution of flux in the reconstructed image, which may potentially result in spurious detection. We then calculate the 25% brightest pixel value I_{25} within the primary clump segmentation, and impose another criteria:

$$I_{25} >= I_s + 8\sigma_s \tag{4.2}$$

to (1) guarantee a minimum false positive rate and (2) decrease the probability of blending between the clump segmentations.

(iiv) Finally, a minimal detection area is applied and determined from a linear length scale equal to 3 times the FWHM of the corresponding PSF function.

Based on the above algorithm, we have identified 193 clumps in total from 98 giant arcs. Figure 4.2 lists some representative detected clumps along with the reconstructed images of lensed galaxies in the source plane. The full image catalog of detected clumps can be found in Figure H.1 in Appendix H

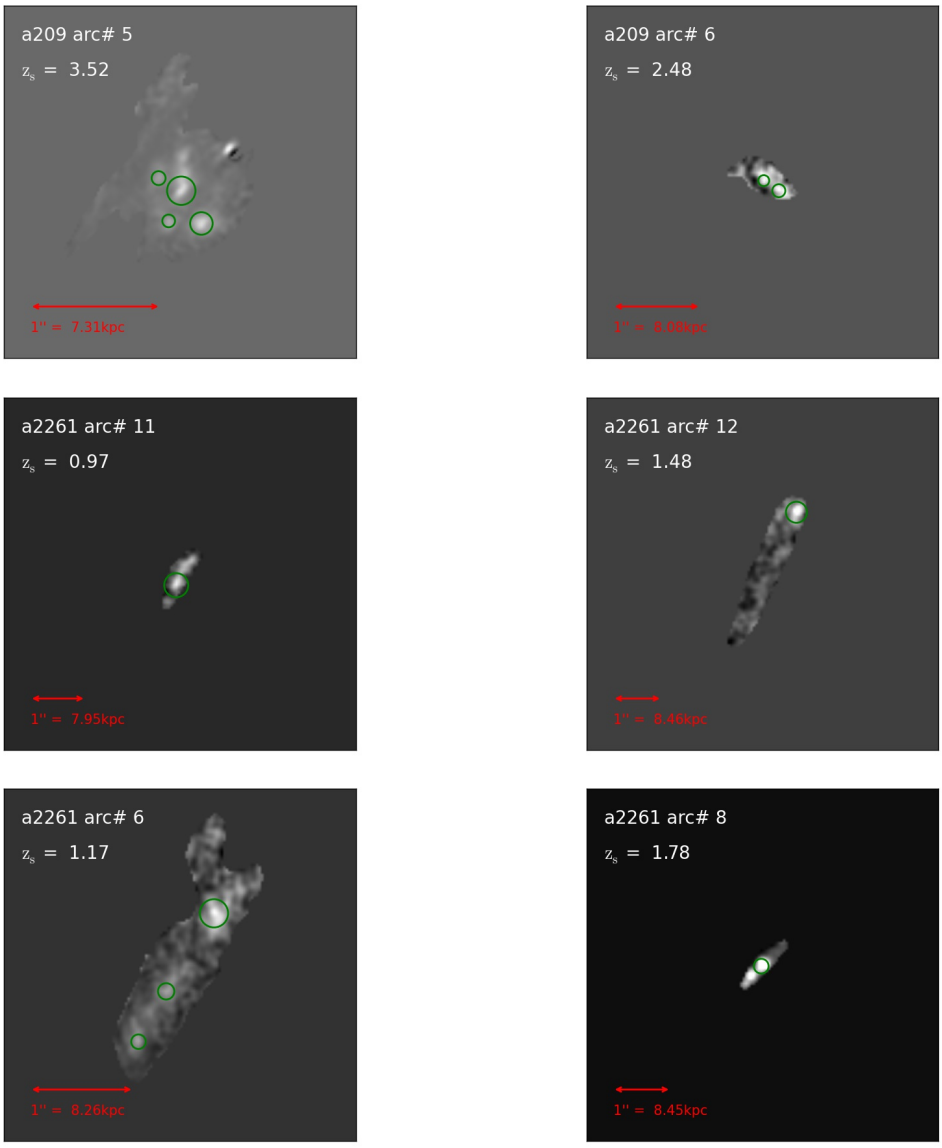


Figure 4.2: The reconstructed rest-frame 2200 $\mathring{A}$ images of lensed galaxies in the source plane. The green circles mark the size and position of the detected clumps.

4.3 Clump Photometry and SED Fitting

As discussed in Chapter 3, the multi-wavelength photometry needs to be extracted using matched PSFs. Unfortunately, a direct PSF-matched photometry measurement is hard to perform in the source plane due to the lensing distortion of the shape of the PSF function. Thus we choose to measure the photometry in the image plane instead. To achieve that, for each detected clump, we define its corresponding aperture in the image plane by re-lensing the clump segmentation map in the source plane back to the image plane via the CLASH mass models. Then the PSF-matched photometry is measured within the generated aperture in the image plane. In extracting the photometry, we do not subtract the background diffuse flux from the host galaxy, since the necessity of subtracting the diffuse background is still an open issue in the clump studies (Forster Schreiber et al., 2011; Guo et al., 2012; Wuyts et al., 2012; Guo et al., 2015). We compute the H-band flux-weighted magnification as the overall magnification factor for each clump and correct the photometry for the lensing magnification. We adopt the same parameters as discussed in Section 3.6.2. to run the SED fitting for clumps.

To investigate the rest-frame UV luminosity function and color of the clumps, we in specific compute the rest-frame AB absolute magnitude in two wavelength: 2200 $\mathring{A}$ and 6500 $\mathring{A}$. For each clump, we use iSEDfit to generate 10000

model spectra by randomly sampling the model parameter space. By assuming an uniform prior, each model parameter will receive a weight that is proportional to its posterior probability. For each model spectra, we then calculate the value of M_{2200} and M_{6500} by computing the average value in a wavelength interval with width of 100 $\mathring{A}$ centered at 2200 $\mathring{A}$ and 6500 $\mathring{A}$, respectively. The posterior distribution of M_{2200} and M_{6500} are then calculated by weighting the computed M_{2200} and M_{6500} values by the weight of the model parameter. The final value and uncertainty of the rest-frame M_{2200} and M_{6500} are adopted as the mode value and 1σ standard deviation of their posterior distributions.

4.4 Size of Clumps

4.4.1 Measuring the Clump Size

Investigating the surface mass density and surface star formation rate density and their evolutions requires accurate size measurement of clumps. There are two commonly adopted methods to measure the size of clumps in literature, the isophotal method and the core method in which a 2D parametric profile is fit to the light profile of clumps (e.g. Wisnioski et al. (2012); Wuyts et al. (2014)). The isophotal radius is determined by a chosen surface brightness threshold which is often chosen arbitrarily and makes the comparison between

the studies difficult. Moreover, the cosmological surface brightness dimming effect renders the comparison of the size measurement at different redshift based on isophotal method difficult. Therefore, we adopt the more robust profile fitting method in this study to compute the size of clumps. For each clump, we compute the distance of each pixel that belongs to the clump to the center of clump. We then divide the pixel distance into 5 bins. In each bin, a median value of pixel intensity and 1σ standard deviation are computed. We then use a parametric function to fit the surface brightness profile of clumps. For simplicity, we use 1D Gaussian profile for the fitting and adopt the FWHM of the best fitted distribution as the size of the clump.

4.4.2 The PSF Correction for the Size Measurement

The PSF blurring makes the image of galaxy or clump larger and puffier, which can be an severe problem for the size measurement, especially for when the physical clump size is less than that corresponding to 4 $\times$ the FWHM of the PSF. This bias becomes larger for small objects at high redshift with comparable angular size with FWHM of the PSF function. Therefore, a reliable size measurement needs to account for the PSF blurring effect. For galaxies or clumps in field, one can generally use the forward fitting method provided by

the software such as GALFIT (Peng et al., 2002) to accurately determine the true size of the objects. Things become more complicated for the lensing selected objects because the shape of the PSF function is distorted in the source plane by the strong lensing effect. In principle, the forward fitting method still applies in determining the true size of the lensed objects. However, that requires the object to be re-lensed from source plane to image plane in each iteration, which adds computational complexity. So alternatively, we consider to use simulated image sets to statistically assess the bias and error from the lensing magnification and to correct the size measurement.

The simulations are performed as follows: first, we simulate sets of template galaxies and clumps at $z = 2$ with 10 different intrinsic half-light radius: r_h =0.5,0.6,0.7,0.8,1.0,1.5,2.0,2.5,3,4 kpc at a resolution = 0".005 per pixel. We are only focusing on how the effects such as lensing magnification, pixel binning, PSF blurring and sky noise affect the size measurement, so we normalize the average surface brightness of all the simulated objects to a constant level which equals the average surface brightness of a galaxy at $z = 2$ with half light radius r_h =1 kpc and absolute AB magnitude $M = -21$ at HST F606W filter [1]. To account for different spatial resolutions, we bin the images down to pixel scale of 0.065", 0.033", 0.022" and 0.016" per pixel, which approximately correspond to $\times$1, $\times$2, $\times$3 and $\times$4 resolution of CLASH images. We place these galax-

[1]Due to the lack of rest-frame optical images for clumps at high redshift, we perform the size measurement for clumps at rest-frame 2200$\mathring{A}$ for consistency.

ies with different pixel scales around the caustic line at different magnification at $z = 2$, and use ray-tracing technique to create the simulated lensed images in the image plane, and mimic the observation procedures to convolve the image with the PSF function of HST F606W filter, bin the images down to the pixel scale of 0.”065 and add Gaussian noise that matches that of CLASH images. We then perform the same image reconstruction procedure on these simulated images as we do to the CLASH data, as discussed in Chapter 3. Given the simulated reconstructed images in the source plane, we use the same core method to measure the size of the simulated reconstructed images and compare with their true sizes. Figure 4.3 shows the ratio of the observed size to the true size as a function of the true size at 4 different magnifications. The trend is consistent with our expectation: as the magnification increases, the size measurement becomes more accurate, especially at small radius end. For objects at sub-kpc scale, the fractional bias at highest magnification remains $\sim 10\%$; whereas the size can be over-estimated by a factor of 50% - 100% without magnification. Hence, though the strong lensing effect adds more complexity on the size measurement, it also significantly increases the spatial resolution and improve the accuracy of size measurement for the small distant objects. This allows the study at spatial scale that are impossible to achieve for field galaxies.

Finally, we categorize the measured size of clumps into four linear magnifi-

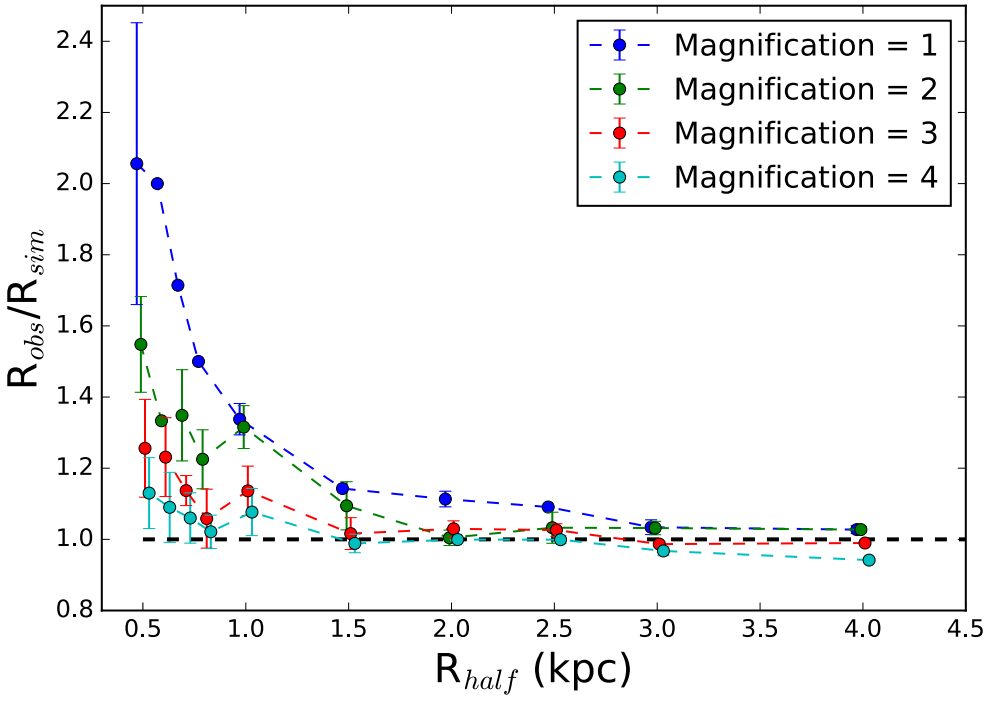


Figure 4.3: The ratio of the measured clump size to the simulated clump size as a function of the simulated size for 4 different linear magnifications. The error bar denotes 1σ rms error.

cation bins by their magnification factors, and correct their size for the magnification bias based on the results of the above simulations.

Table 4.1 lists the measured physical properties of the all detected clumps.

Table 4.1: Detected clumps and their physical properties.

Cluster	ID Arc	ID Clump	z	M_{2200}	M_{6500}	SFR $\log(M_\odot\ yr^{-1})$	MSTAR $\log(M_\odot)$	sSFR $\log(\mathrm{Gyr}^{-1})$	Age Gyr	rad kpc	d_{proj}/R_e
a209	6	1	2.48	-17.06±0.20	-17.38±0.14	0.73±0.13	7.70±0.05	2.03±0.14	0.01±0.01	0.60	7.3
a209	6	2	2.48	-17.48±0.39	-18.12±0.33	1.02±0.17	8.13±0.11	1.90±0.21	0.02±0.01	0.51	0.3
a2261	12	1	1.48	-18.47±0.39	-20.22±0.37	1.81±0.29	9.27±0.11	1.54±0.31	0.05±0.06	1.83	27.2
a2261	6	1	1.17	-16.39±0.05	-17.21±0.04	0.07±0.00	8.19±0.01	0.88±0.01	0.15±0.00	0.58	35.8
a2261	6	2	1.17	-16.38±0.09	-18.47±0.11	0.27±0.09	9.20±0.03	0.07±0.10	0.67±0.12	0.65	12.9
a2261	6	3	1.17	-16.77±0.27	-17.50±0.28	0.14±0.17	8.32±0.08	0.82±0.18	0.20±0.09	1.13	23.4
a2261	9	1	3.09	-19.47±0.26	-21.39±0.16	1.19±0.06	10.51±0.09	-0.32±0.11	1.44±0.29	1.09	9.8
a383	12	1	3.17	-17.31±0.77	-17.13±0.49	-0.23±0.12	7.77±0.28	0.99±0.30	0.15±0.12	0.44	24.5
a383	12	2	3.17	-15.57±0.99	-15.75±0.63	-0.55±0.20	7.46±0.34	0.99±0.39	0.18±0.19	0.88	10.0
a383	3	1	3.96	-17.60±1.36	-17.23±0.83	-0.24±0.17	7.62±0.47	1.14±0.50	0.20±0.32	0.36	46.5
a383	3	2	3.96	-16.96±1.16	-17.65±0.68	-0.03±0.13	8.53±0.43	0.44±0.45	0.51±0.34	0.60	11.5
a383	3	3	3.96	-15.98±1.15	-16.67±0.68	-0.48±0.20	8.15±0.41	0.37±0.46	0.55±0.36	0.41	8.4
a383	3	4	3.96	-17.22±1.00	-17.19±0.60	-0.32±0.17	8.30±0.37	0.39±0.41	0.52±0.35	0.40	21.3

To be continued

Cluster	ID Arc	ID Clump	z	M_{2200}	M_{6500}	SFR $\log(M_\odot\ yr^{-1})$	MSTAR $\log(M_\odot)$	sSFR $\log(\mathrm{Gyr}^{-1})$	Age Gyr	rad kpc	d_{proj}/R_e
a383	6	1	2.73	-17.23±0.26	-18.42±0.21	0.37±0.14	8.56±0.07	0.82±0.16	0.17±0.07	0.42	5.5
a383	6	2	2.73	-17.69±0.23	-18.72±0.22	0.87±0.12	8.67±0.09	1.21±0.16	0.07±0.03	0.22	30.7
a383	6	3	2.73	-19.08±0.08	-19.27±0.05	0.63±0.03	8.82±0.01	0.82±0.03	0.16±0.02	0.24	26.8
a383	9	1	2.64	-18.29±0.44	-19.06±0.29	0.20±0.13	9.17±0.16	0.03±0.21	0.82±0.35	0.33	18.0
a383	9	2	2.64	-18.07±0.45	-19.43±0.29	0.18±0.14	9.44±0.16	-0.26±0.21	1.24±0.47	0.90	9.0
a383	9	3	2.64	-18.15±0.39	-18.78±0.24	0.09±0.09	8.87±0.15	0.22±0.17	0.56±0.19	0.81	24.2
a611	1	1	2.44	-18.61±0.24	-20.22±0.18	1.55±0.15	9.52±0.06	1.03±0.16	0.10±0.03	0.56	15.3
a611	3	1	1.00	-14.78±0.04	-16.49±0.04	-1.25±0.00	8.18±0.02	-0.43±0.02	0.48±0.06	0.44	0.4
a611	3	2	1.00	-15.65±0.04	-17.27±0.05	-0.85±0.04	8.66±0.02	-0.51±0.05	1.66±0.18	0.84	21.7
clj1226	10	1	1.15	-15.54±0.21	-14.87±0.24	-0.70±0.11	6.82±0.08	1.48±0.14	0.04±0.01	0.87	4.8
clj1226	10	2	1.15	-15.10±0.42	-17.60±0.17	0.86±0.00	8.56±0.00	1.30±0.01	0.05±0.00	1.06	4.5
clj1226	2	1	3.55	-15.73±0.12	-14.98±0.09	-0.56±0.01	6.39±0.01	2.05±0.02	0.01±0.00	1.01	7.6
clj1226	2	2	3.55	-12.32±1.19	-15.85±0.62	-0.50±0.49	8.54±0.44	-0.04±0.65	0.80±0.41	1.24	7.5
clj1226	6	1	2.61	-18.78±0.46	-18.64±0.40	0.56±0.19	8.13±0.17	1.43±0.26	0.06±0.04	0.40	9.6
clj1226	6	2	2.61	-17.94±0.51	-18.32±0.40	0.35±0.21	8.33±0.17	1.02±0.27	0.13±0.07	0.65	12.5
clj1226	6	3	2.61	-18.95±0.21	-19.29±0.17	0.55±0.08	8.66±0.07	0.89±0.10	0.12±0.04	2.02	7.9
macs0329	1	1	3.48	-20.30±0.29	-21.99±0.19	1.15±0.13	10.68±0.11	-0.53±0.17	1.35±0.25	1.51	18.9
macs0329	2	1	5.77	-17.98±1.09	-19.65±0.58	0.82±0.34	9.50±0.42	0.31±0.54	0.48±0.23	0.48	3.3
macs0329	2	2	5.77	-18.45±1.02	-19.57±0.55	0.73±0.26	9.41±0.41	0.32±0.48	0.47±0.22	0.23	29.9
macs0329	3	1	6.19	-18.14±1.16	-18.48±0.66	0.30±0.22	8.78±0.48	0.52±0.53	0.37±0.21	1.01	28.1
macs0329	3	2	6.19	-17.81±1.23	-18.25±0.70	0.40±0.23	8.83±0.49	0.58±0.55	0.35±0.21	0.90	2.5
macs0329	3	3	6.19	-17.19±1.36	-18.09±0.74	0.25±0.32	8.72±0.58	0.53±0.67	0.37±0.20	1.23	9.8
macs0329	5	1	6.19	-17.99±1.08	-18.96±0.61	0.51±0.27	9.07±0.43	0.44±0.51	0.41±0.21	1.89	1.7
macs0329	5	2	6.19	-17.89±1.26	-18.54±0.74	0.55±0.21	8.91±0.48	0.64±0.52	0.33±0.21	0.51	24.2
macs0329	6	1	1.38	-15.27±0.34	-16.74±0.29	-0.49±0.37	8.29±0.10	0.22±0.38	0.59±0.33	0.55	12.9
macs0329	6	2	1.38	-15.51±0.29	-16.73±0.24	-0.60±0.28	8.29±0.10	0.11±0.29	0.71±0.40	0.31	25.5
macs0329	6	3	1.38	-16.38±0.15	-17.34±0.16	-0.32±0.21	8.29±0.05	0.38±0.22	0.21±0.12	0.37	9.9
macs0329	7	1	2.77	-18.32±0.34	-18.51±0.24	0.25±0.12	8.70±0.12	0.54±0.17	0.29±0.13	0.23	2.3
macs0329	7	2	2.77	-18.08±0.34	-18.27±0.25	0.19±0.14	8.57±0.13	0.62±0.19	0.25±0.11	0.43	12.4
macs0429	2	1	1.50	-15.98±0.64	-17.14±0.44	-0.19±0.35	8.43±0.18	0.38±0.40	0.52±0.36	0.46	1.3
macs0429	2	2	1.50	-18.32±0.18	-20.08±0.16	0.56±0.22	9.80±0.07	-0.24±0.23	1.16±0.45	0.87	15.5
macs0429	4	1	2.42	-18.87±0.32	-19.93±0.28	1.16±0.20	9.54±0.10	0.62±0.22	0.27±0.16	0.85	14.7
macs0429	4	2	2.42	-17.19±0.72	-18.86±0.45	0.22±0.35	9.26±0.23	-0.04±0.42	1.01±0.60	1.04	45.4
macs0429	7	1	4.06	-18.65±0.97	-18.97±0.60	0.37±0.16	8.74±0.34	0.63±0.38	0.35±0.25	0.89	8.9
macs0429	7	2	4.06	-18.97±0.68	-19.90±0.39	0.52±0.11	9.48±0.25	0.04±0.27	0.78±0.33	0.82	1.7
macs0429	7	3	4.06	-19.08±0.83	-19.42±0.51	0.55±0.15	9.09±0.32	0.45±0.35	0.42±0.29	1.08	18.6
macs0429	8	1	1.14	-16.30±0.21	-18.41±0.20	0.30±0.30	9.16±0.07	0.14±0.30	0.70±0.46	1.59	0.7
macs0744	2	1	4.88	-19.60±1.01	-20.72±0.60	1.13±0.15	9.85±0.37	0.28±0.40	0.58±0.28	0.99	12.4
macs0744	2	2	4.88	-19.22±1.16	-19.15±0.70	0.59±0.17	8.71±0.41	0.88±0.45	0.25±0.25	0.85	12.3
macs0744	2	3	4.88	-17.08±1.23	-18.17±0.71	0.30±0.20	8.87±0.45	0.44±0.49	0.48±0.28	0.54	27.1
macs0744	6	1	4.03	-18.96±1.12	-17.57±0.67	0.06±0.14	8.30±0.39	0.76±0.41	0.29±0.26	0.63	14.4
macs0744	7	1	3.96	-18.37±0.75	-18.14±0.53	0.25±0.19	8.19±0.27	1.06±0.33	0.14±0.12	0.55	4.1
macs0744	7	2	3.96	-18.61±0.47	-19.09±0.26	0.28±0.06	9.19±0.18	0.08±0.19	0.75±0.25	0.98	2.1
macs0744	9	1	4.00	-16.69±1.00	-17.23±0.58	-0.16±0.24	8.16±0.38	0.68±0.45	0.33±0.29	1.47	3.4
macs1115	11	1	3.14	-18.01±0.59	-20.29±0.39	1.01±0.25	10.15±0.20	-0.14±0.32	1.10±0.46	0.59	20.2
macs1206	10	1	1.51	-15.45±0.39	-17.96±0.26	-0.21±0.56	9.23±0.10	-0.44±0.57	1.45±0.88	0.57	12.0
macs1206	10	2	1.51	-18.03±0.18	-19.82±0.15	1.48±0.05	9.33±0.04	1.15±0.06	0.08±0.01	0.36	2.1
macs1206	10	3	1.51	-15.20±0.50	-17.52±0.40	-0.13±0.49	9.00±0.14	-0.13±0.51	1.28±0.82	0.65	4.2
macs1206	11	1	2.72	-18.69±0.17	-19.89±0.17	1.20±0.09	9.42±0.06	0.79±0.11	0.18±0.05	0.31	18.6
macs1206	11	2	2.72	-18.43±0.19	-19.48±0.14	1.70±0.13	8.76±0.08	1.95±0.15	0.01±0.01	0.59	8.1
macs1206	4	1	1.29	-14.06±2.53	-14.76±1.18	-1.64±0.80	7.16±0.51	0.20±0.95	0.78±1.01	0.63	10.4
macs1206	4	2	1.29	-15.73±0.32	-16.96±0.24	-0.68±0.22	8.42±0.10	-0.10±0.24	1.03±0.38	0.82	0.3
macs1206	5	1	2.33	-15.21±0.76	-16.55±0.48	-0.59±0.27	8.37±0.25	0.05±0.37	0.88±0.56	0.35	29.7
macs1206	5	2	2.33	-15.57±0.59	-16.86±0.36	-0.77±0.22	8.50±0.20	-0.27±0.30	1.30±0.59	0.58	10.2
macs1206	7	1	3.05	-19.15±0.38	-20.39±0.25	0.65±0.10	9.71±0.14	-0.06±0.17	0.95±0.32	0.43	11.2
macs1206	8	1	2.07	-17.90±0.27	-18.40±0.26	0.43±0.17	8.64±0.08	0.79±0.19	0.19±0.08	0.29	15.6
macs1206	8	2	2.07	-13.45±2.10	-14.95±1.01	-1.31±0.45	7.63±0.59	0.06±0.74	0.91±0.72	0.48	3.8
macs1206	8	3	2.07	-15.94±0.55	-16.63±0.45	-0.15±0.31	8.04±0.18	0.81±0.36	0.23±0.18	0.30	18.0
macs1311	1	1	1.18	-12.78±3.30	-13.13±1.36	-2.49±1.50	7.40±0.54	-0.89±1.60	1.87±1.22	0.65	15.0
macs1311	1	2	1.18	-15.43±0.41	-15.91±0.36	-0.56±0.26	7.60±0.13	0.84±0.29	0.19±0.12	0.83	39.2
macs1311	2	1	0.99	-16.66±0.16	-18.25±0.15	-0.24±0.13	9.08±0.03	-0.33±0.14	1.44±0.42	1.44	17.1
macs1311	6	1	0.92	-15.10±0.70	-17.57±0.36	-0.68±0.58	8.91±0.14	-0.59±0.60	1.70±1.07	0.64	13.1
macs1311	6	2	0.92	-16.65±0.33	-18.30±0.33	0.19±0.40	9.12±0.10	0.07±0.41	0.92±0.56	0.68	20.6
macs1311	6	3	0.92	-15.78±0.46	-17.18±0.37	-0.28±0.35	8.54±0.14	0.18±0.38	0.73±0.42	0.47	3.5
macs1423	10	1	1.80	-18.64±0.15	-19.87±0.14	0.78±0.10	9.44±0.04	0.34±0.11	0.42±0.10	0.81	23.6
macs1423	5	1	1.45	-15.22±0.25	-17.71±0.21	-0.19±0.47	9.14±0.09	-0.33±0.48	1.21±0.82	0.73	6.6
macs1423	6	1	2.17	-16.78±0.22	-17.23±0.21	-0.37±0.09	8.14±0.07	0.48±0.11	0.34±0.10	0.39	0.6
macs1423	6	2	2.17	-17.21±0.15	-17.53±0.09	0.66±0.00	7.82±0.00	1.84±0.00	0.02±0.00	0.39	6.1
macs1423	6	3	2.17	-16.10±0.43	-16.46±0.41	-0.23±0.25	7.73±0.14	1.04±0.29	0.13±0.09	0.36	33.9
macs1423	6	4	2.17	-17.56±0.31	-17.94±0.31	0.53±0.12	8.01±0.09	1.52±0.14	0.03±0.02	0.34	25.6
macs1720	1	1	2.71	-19.19±0.00	-20.38±0.00	0.37±0.00	9.67±0.00	-0.29±0.00	0.45±0.00	0.70	14.6
macs1720	2	1	4.43	-16.91±1.34	-18.14±0.76	0.27±0.21	8.89±0.54	0.39±0.58	0.53±0.32	0.60	9.6
macs1720	2	2	4.43	-19.09±1.39	-19.07±0.85	0.58±0.14	8.63±0.51	0.95±0.53	0.24±0.29	0.54	19.6
macs1720	2	3	4.43	-17.54±1.47	-18.50±0.74	0.26±0.39	9.03±0.61	0.23±0.72	0.60±0.32	0.96	32.4
macs1931	1	1	3.77	-19.58±0.92	-21.61±0.56	1.78±0.30	10.50±0.34	0.28±0.45	0.57±0.37	1.01	2.6
macs1931	2	1	2.42	-17.61±0.25	-18.35±0.25	0.63±0.14	8.70±0.08	0.93±0.16	0.14±0.05	1.05	11.5
macs1931	2	2	2.42	-16.90±0.34	-17.91±0.30	0.62±0.29	8.40±0.10	1.22±0.30	0.08±0.05	0.58	20.8
macs1931	3	1	4.54	-18.32±0.95	-19.29±0.56	0.60±0.17	9.34±0.35	0.26±0.39	0.59±0.30	0.53	9.0
macs1931	3	2	4.54	-17.10±1.12	-18.15±0.66	0.20±0.20	8.85±0.39	0.36±0.44	0.55±0.31	0.86	23.8
macs1931	4	1	3.21	-18.12±0.63	-18.91±0.47	0.66±0.24	8.63±0.21	1.03±0.32	0.14±0.11	1.10	13.0
macs1931	4	2	3.21	-18.02±0.56	-18.49±0.42	0.50±0.16	8.70±0.22	0.80±0.27	0.20±0.09	0.88	2.0
ms2137	1	1	1.38	-17.06±0.33	-17.70±0.22	-0.22±0.19	8.52±0.07	0.26±0.20	0.48±0.21	1.36	8.4
rxj1347	1	1	1.37	-13.53±0.57	-16.32±0.37	-1.19±1.50	8.33±0.15	-0.52±1.51	0.94±0.95	1.50	1.4

To be continued

Cluster	ID Arc	ID Clump	z	M_{2200}	M_{6500}	SFR $\log(M_\odot\ yr^{-1})$	MSTAR $\log(M_\odot)$	sSFR $\log(\mathrm{Gyr}^{-1})$	Age Gyr	rad kpc	d_{proj}/R_e
rxj1347	1	2	1.37	-14.05±0.42	-16.73±0.35	0.27±0.31	8.41±0.12	0.86±0.33	0.20±0.24	1.04	6.8
rxj1347	1	3	1.37	-15.90±0.42	-18.08±0.42	0.41±0.39	8.75±0.15	0.66±0.42	0.37±0.35	1.07	7.4
rxj1347	1	4	1.37	-13.32±0.62	-15.56±0.43	-1.06±0.48	8.16±0.18	-0.22±0.51	1.38±0.88	0.62	9.5
rxj1347	2	1	1.76	-18.23±0.04	-18.84±0.02	0.68±0.00	8.79±0.00	0.89±0.00	0.13±0.00	0.53	24.6
rxj1347	2	2	1.76	-17.80±0.09	-18.48±0.09	0.48±0.04	8.69±0.02	0.78±0.05	0.17±0.02	0.56	4.9
rxj1347	2	3	1.76	-16.85±0.05	-17.65±0.05	0.15±0.06	8.41±0.02	0.74±0.07	0.19±0.03	0.52	10.1
rxj1347	6	1	4.18	-17.64±0.77	-17.27±0.49	-0.12±0.16	7.73±0.29	1.15±0.34	0.11±0.14	0.78	17.1
rxj1347	6	2	4.18	-17.31±1.01	-17.50±0.60	-0.02±0.16	8.60±0.37	0.39±0.41	0.52±0.34	1.28	1.1
rxj1347	8	1	3.57	-18.02±0.84	-17.64±0.53	0.06±0.14	8.35±0.32	0.70±0.35	0.28±0.19	0.35	7.9
rxj1347	9	1	3.01	-18.84±0.15	-20.32±0.10	0.90±0.03	9.86±0.07	0.03±0.08	0.77±0.13	0.82	2.4
rxj2129	2	1	2.95	-19.89±0.42	-21.95±0.38	2.42±0.30	10.45±0.16	0.97±0.34	0.13±0.09	0.39	14.7
rxj2129	2	2	2.95	-19.00±0.56	-21.23±0.41	1.65±0.28	10.12±0.22	0.53±0.36	0.40±0.38	0.82	6.8
rxj2248	10	1	1.46	-18.86±0.10	-20.28±0.09	0.82±0.07	9.65±0.02	0.18±0.07	0.33±0.04	0.65	10.6
rxj2248	10	2	1.46	-16.27±0.40	-17.00±0.28	-0.46±0.19	8.11±0.12	0.43±0.22	0.39±0.17	0.54	9.3
rxj2248	11	1	0.91	-13.32±3.55	-14.37±1.50	-2.48±1.50	7.14±0.52	-0.62±1.59	1.68±1.43	0.73	3.2
rxj2248	12	1	1.36	-15.55±0.39	-16.03±0.32	-0.67±0.27	7.74±0.11	0.59±0.29	0.30±0.16	0.56	9.8
rxj2248	14	1	2.79	-17.43±0.18	-17.60±0.07	-0.12±0.00	8.26±0.00	0.62±0.00	0.24±0.00	0.74	3.6
rxj2248	14	2	2.79	-17.43±0.06	-18.49±0.03	1.64±0.00	8.43±0.00	2.21±0.00	0.01±0.00	1.00	10.3
rxj2248	15	1	4.34	-16.78±1.17	-18.33±0.66	0.02±0.21	8.76±0.46	0.26±0.51	0.61±0.32	1.35	2.5
rxj2248	16	1	2.07	-14.71±0.33	-15.99±0.28	0.23±0.26	7.47±0.09	1.76±0.28	0.02±0.02	0.72	20.1
rxj2248	16	2	2.07	-14.17±0.30	-15.04±0.22	-1.28±0.24	7.65±0.09	0.07±0.26	0.70±0.32	0.48	10.0
rxj2248	16	3	2.07	-12.21±0.94	-13.82±0.46	-2.07±0.22	7.41±0.26	-0.48±0.34	1.71±0.63	0.62	10.6
rxj2248	16	4	2.07	-14.87±0.32	-15.43±0.20	-1.15±0.10	7.56±0.08	0.29±0.13	0.49±0.14	0.45	29.3
rxj2248	2	1	2.21	-17.90±0.27	-17.41±0.21	0.07±0.14	8.01±0.08	1.06±0.17	0.10±0.03	0.42	6.3
rxj2248	5	1	3.40	-19.52±0.26	-21.49±0.16	0.73±0.07	10.29±0.11	-0.56±0.13	1.44±0.20	0.82	16.1
a209	5	1	3.52	-19.09±0.67	-20.24±0.42	1.00±0.20	9.65±0.27	0.35±0.34	0.46±0.29	0.41	31.1
a209	5	2	3.52	-18.59±0.45	-19.56±0.23	0.31±0.03	9.38±0.17	-0.07±0.18	0.99±0.30	0.45	10.5
a209	5	3	3.52	-20.25±0.41	-21.74±0.28	0.92±0.12	10.21±0.14	-0.29±0.18	0.93±0.32	0.39	5.1
a209	5	4	3.52	-17.92±0.75	-19.74±0.42	0.49±0.16	9.65±0.30	-0.16±0.34	1.03±0.39	0.36	13.9
a611	2	1	3.00	-20.23±0.00	-19.61±0.00	1.22±0.00	8.30±0.00	1.92±0.00	0.01±0.00	0.39	8.8
a611	2	2	3.00	-21.36±0.07	-21.47±0.03	2.00±0.02	9.43±0.02	1.57±0.03	0.03±0.00	0.40	12.2
a611	2	3	3.00	-20.91±0.15	-20.83±0.05	1.24±0.01	9.40±0.01	0.84±0.02	0.15±0.00	0.63	9.9
a611	2	4	3.00	-19.96±0.16	-19.20±0.10	1.28±0.09	8.13±0.05	2.15±0.10	0.01±0.00	0.90	2.3
a2261	8	1	1.78	-19.17±0.08	-20.49±0.04	1.27±0.05	9.73±0.01	0.54±0.05	0.27±0.00	1.70	14.3
a2261	11	1	0.97	-18.17±0.10	-19.44±0.09	0.35±0.09	9.41±0.03	-0.06±0.09	0.85±0.16	1.03	9.2
clj1226	3	1	2.76	-19.31±0.25	-19.98±0.17	0.56±0.08	9.36±0.09	0.20±0.12	0.59±0.20	0.81	0.7
clj1226	5	1	3.57	-14.05±1.33	-15.60±0.76	-0.96±0.22	7.67±0.48	0.37±0.53	0.57±0.39	1.21	7.7
clj1226	5	2	3.57	-14.04±1.17	-17.52±0.67	0.26±0.31	9.05±0.43	0.21±0.53	0.62±0.40	0.74	42.6
clj1226	7	1	4.15	-19.15±0.86	-20.37±0.54	1.19±0.25	9.59±0.29	0.61±0.38	0.34±0.27	0.69	16.2
clj1226	7	2	4.15	-15.96±1.47	-18.25±0.71	0.26±0.50	9.28±0.56	-0.02±0.75	0.74±0.34	0.54	11.1
clj1226	8	1	2.80	-18.39±0.65	-19.88±0.41	0.54±0.23	9.58±0.23	-0.04±0.32	0.94±0.49	0.84	31.5
clj1226	9	1	1.59	-15.41±0.77	-16.78±0.48	-0.29±0.43	8.35±0.20	0.35±0.48	0.58±0.51	0.68	44.5
clj1226	9	2	1.59	-17.59±0.26	-18.31±0.21	0.07±0.15	8.75±0.09	0.31±0.17	0.47±0.18	0.47	7.0
macs0329	4	1	2.88	-15.66±0.26	-14.71±0.17	-0.70±0.02	6.50±0.11	1.80±0.11	0.02±0.01	0.38	15.5
macs0429	1	1	2.65	-18.90±0.36	-19.88±0.27	0.58±0.11	9.56±0.13	0.02±0.17	0.81±0.33	0.28	6.4
macs0744	8	1	0.94	-15.72±0.23	-16.31±0.21	-0.23±0.13	7.51±0.08	1.27±0.15	0.06±0.03	0.31	17.1
macs0744	8	2	0.94	-15.34±0.24	-16.67±0.24	0.20±0.15	7.99±0.08	1.21±0.18	0.07±0.02	0.51	16.1
macs0744	8	3	0.94	-15.25±0.16	-16.79±0.16	0.31±0.07	7.96±0.05	1.35±0.08	0.05±0.02	0.81	6.2
macs0744	14	1	2.83	-18.15±0.07	-19.94±0.04	-0.30±0.00	9.52±0.00	-0.82±0.00	0.32±0.00	0.89	4.3
macs0744	14	2	2.83	-18.37±0.21	-20.02±0.08	-0.17±0.08	9.53±0.05	-0.70±0.10	0.32±0.00	0.42	8.3
macs0744	14	3	2.83	-17.48±0.46	-18.18±0.28	-0.14±0.07	8.69±0.15	0.17±0.17	0.62±0.23	0.41	2.8
macs1115	8	1	0.77	-14.92±0.36	-17.86±0.18	-0.99±0.27	9.37±0.06	-1.36±0.28	3.54±0.98	0.92	15.5
macs1115	8	2	0.77	-14.53±0.24	-17.64±0.16	-0.79±0.26	9.21±0.06	-1.01±0.27	2.83±0.92	0.52	28.3
macs1311	3	1	0.72	-14.26±1.83	-14.57±0.95	-1.45±0.51	7.00±0.37	0.55±0.63	0.50±0.71	0.65	36.3
macs1311	3	2	0.72	-15.05±0.32	-16.79±0.20	-1.00±0.46	8.39±0.07	-0.39±0.46	0.91±0.42	0.63	19.1
macs1423	2	1	2.82	-18.12±0.22	-19.32±0.16	0.44±0.08	9.36±0.10	0.08±0.13	0.67±0.23	0.36	17.6
macs1423	9	1	1.75	-19.21±0.03	-20.41±0.01	1.42±0.08	9.63±0.01	0.79±0.08	0.17±0.04	0.80	4.7
macs1423	11	1	3.17	-20.37±0.43	-20.72±0.29	0.99±0.11	9.56±0.15	0.42±0.19	0.38±0.19	0.39	9.7
macs1423	11	2	3.17	-18.12±0.87	-19.08±0.52	0.28±0.17	9.13±0.33	0.15±0.37	0.73±0.42	1.21	51.8
macs1423	12	1	3.65	-19.47±0.33	-21.93±0.17	0.81±0.07	10.60±0.13	-0.79±0.14	1.47±0.14	1.05	2.8
macs1931	5	1	0.78	-16.28±0.22	-17.43±0.19	-0.43±0.16	8.57±0.08	-0.00±0.18	0.82±0.28	0.89	15.2
rxj1347	4	1	2.45	-15.42±1.20	-16.10±0.72	-0.70±0.24	7.84±0.47	0.47±0.53	0.51±0.46	1.08	56.4
rxj1347	4	2	2.45	-16.55±0.94	-17.03±0.64	-0.10±0.29	8.19±0.34	0.71±0.45	0.32±0.27	1.08	2.1
rxj1532	2	1	2.91	-19.31±0.35	-20.38±0.23	0.77±0.05	9.85±0.13	-0.08±0.14	0.97±0.29	1.74	5.5
rxj1532	2	2	2.91	-20.81±0.16	-21.62±0.14	1.49±0.04	10.12±0.05	0.37±0.07	0.42±0.07	0.85	10.6
rxj1532	3	1	0.64	-16.53±0.07	-18.70±0.13	-0.03±0.17	9.28±0.04	-0.30±0.18	0.88±0.50	0.26	1.8
rxj1532	3	2	0.64	-17.22±0.11	-18.50±0.13	-0.24±0.01	8.99±0.03	-0.23±0.03	0.66±0.20	0.24	40.7
rxj2248	3	1	0.94	-14.58±0.45	-14.41±0.38	-0.91±0.31	6.65±0.14	1.44±0.34	0.06±0.05	1.18	9.5
rxj2248	3	2	0.94	-14.79±0.31	-14.87±0.22	-1.10±0.12	6.96±0.10	0.93±0.16	0.12±0.05	0.31	51.2
rxj2248	3	3	0.94	-13.17±0.62	-13.54±0.40	-1.71±0.28	6.79±0.18	0.50±0.33	0.37±0.23	2.10	2.6
rxj2248	3	4	0.94	-14.93±0.13	-16.65±0.10	-1.06±0.16	8.40±0.03	-0.46±0.16	1.20±0.51	1.29	9.3
rxj2248	4	1	2.42	-18.06±0.13	-17.41±0.08	0.80±0.07	7.57±0.08	2.23±0.11	0.01±0.00	1.39	10.8
rxj2248	4	2	2.42	-18.25±0.23	-18.49±0.18	0.46±0.07	8.49±0.05	0.97±0.09	0.12±0.03	0.49	11.6
rxj2248	18	1	0.91	-16.10±0.27	-17.07±0.19	-0.53±0.23	8.35±0.09	0.12±0.24	0.62±0.23	2.00	16.4
rxj2248	18	2	0.91	-16.09±0.28	-17.08±0.21	-0.49±0.22	8.31±0.09	0.21±0.23	0.58±0.23	0.70	50.2
rxj2248	18	3	0.91	-15.57±0.42	-16.18±0.33	-0.73±0.30	7.95±0.13	0.31±0.33	0.52±0.27	0.50	25.3
a383	4	1	0.85	-16.04±0.12	-17.76±0.11	0.27±0.06	8.66±0.04	0.60±0.07	0.26±0.07	0.55	10.7
a383	4	2	0.85	-18.53±0.00	-19.05±0.00	0.99±0.00	8.81±0.00	1.18±0.00	0.07±0.00	1.18	12.8
a383	4	3	0.85	-16.68±0.14	-18.17±0.11	-0.39±0.15	8.99±0.04	-0.38±0.16	1.55±0.33	1.15	3.0
a383	11	1	4.62	-16.12±1.47	-17.25±0.86	-0.02±0.15	8.47±0.53	0.51±0.55	0.48±0.30	1.21	9.4
a383	11	2	4.62	-20.30±1.22	-20.01±0.74	0.52±0.12	8.89±0.38	0.64±0.39	0.39±0.30	0.82	6.4
a383	11	3	4.62	-16.06±1.30	-16.74±0.75	-0.37±0.20	8.17±0.49	0.46±0.53	0.48±0.30	0.74	7.2
macs0744	4	1	4.78	-17.48±1.16	-17.15±0.71	-0.14±0.16	7.68±0.40	1.18±0.43	0.15±0.17	1.30	22.1

To be continued

Cluster	ID Arc	ID Clump	z	M_{2200}	M_{6500}	SFR $\log(M_\odot\ yr^{-1})$	MSTAR $\log(M_\odot)$	sSFR $\log(\mathrm{Gyr}^{-1})$	Age Gyr	rad kpc	d_{proj}/R_e
macs1115	2	1	2.39	-15.64±0.18	-16.23±0.17	-0.76±0.14	7.88±0.06	0.36±0.15	0.40±0.11	1.23	1.0
macs1115	2	2	2.39	-16.43±0.19	-17.19±0.18	-0.32±0.24	8.08±0.07	0.61±0.25	0.22±0.10	1.05	0.4
macs1115	2	3	2.39	-16.06±0.21	-17.35±0.20	0.20±0.22	8.48±0.07	0.72±0.23	0.22±0.11	1.11	9.1
macs1115	2	4	2.39	-15.71±0.18	-16.27±0.15	-0.78±0.07	7.83±0.05	0.38±0.09	0.41±0.09	1.37	9.1
macs1115	7	1	1.20	-14.34±0.03	-15.48±0.03	0.18±0.00	7.16±0.00	2.02±0.00	0.01±0.00	1.23	16.0
macs1115	7	2	1.20	-14.18±0.13	-14.68±0.09	-0.55±0.04	6.76±0.03	1.69±0.05	0.02±0.00	1.13	1.5
macs1115	9	1	3.19	-17.87±0.40	-19.74±0.26	0.36±0.13	9.76±0.15	-0.40±0.20	1.33±0.37	1.60	9.7
macs1206	6	1	1.06	-16.21±0.12	-18.31±0.13	0.40±0.21	9.13±0.05	0.27±0.21	0.49±0.19	0.32	9.0
macs1206	6	2	1.06	-16.23±0.29	-18.68±0.28	1.18±0.08	8.98±0.07	1.20±0.11	0.08±0.03	0.56	2.9
macs1206	6	3	1.06	-17.00±0.17	-19.34±0.15	1.52±0.03	9.33±0.05	1.19±0.06	0.07±0.02	0.72	13.3
macs1206	6	4	1.06	-16.61±0.10	-19.35±0.10	0.97±0.12	9.57±0.04	0.40±0.12	0.19±0.14	1.21	3.6
macs1206	6	5	1.06	-16.20±0.22	-18.89±0.19	-0.78±0.45	9.65±0.06	-1.42±0.45	4.03±1.03	0.89	15.8
macs1206	6	6	1.06	-16.89±0.24	-19.18±0.23	1.24±0.20	9.38±0.08	0.86±0.22	0.16±0.09	1.50	25.6
ms2137	3	1	1.79	-17.45±0.06	-17.71±0.06	-0.13±0.03	8.28±0.02	0.59±0.04	0.25±0.03	0.62	34.2
ms2137	3	2	1.79	-17.14±0.03	-18.11±0.01	0.08±0.06	8.72±0.00	0.36±0.06	0.34±0.01	0.66	30.2
ms2137	3	3	1.79	-17.34±0.16	-17.63±0.15	0.31±0.03	7.98±0.03	1.33±0.04	0.05±0.01	0.41	12.3
rxj2129	4	1	1.80	-19.05±0.08	-19.89±0.10	0.86±0.07	9.28±0.05	0.58±0.09	0.26±0.09	0.78	9.5
rxj2248	6	1	1.29	-15.66±0.40	-16.47±0.30	-0.53±0.22	7.93±0.14	0.54±0.26	0.33±0.16	0.92	21.0
rxj2248	6	2	1.29	-16.44±0.22	-17.76±0.19	-0.03±0.23	8.62±0.07	0.35±0.24	0.45±0.22	0.78	29.5
rxj2248	8	1	1.26	-17.80±0.08	-18.55±0.08	0.49±0.15	8.71±0.04	0.78±0.15	0.17±0.03	0.86	6.8
rxj2248	8	2	1.26	-17.06±0.28	-17.89±0.23	0.36±0.21	8.46±0.08	0.89±0.22	0.15±0.07	0.79	4.9
rxj2248	8	3	1.26	-17.62±0.17	-18.32±0.18	0.14±0.27	8.63±0.05	0.51±0.28	0.20±0.08	0.81	13.1

4.5 Completeness Estimation of Clump Detection

There are always some objects missing in galaxy survey due to their faintness. This deficit of faint objects must be corrected when it comes to understanding the luminosity function or stellar mass function that extends to low luminosity or mass end. In this section, we describe the methodology to evaluate the completeness of our clump detection.

The completeness of the clump detection is a complex function of many variables such as clump apparent luminosity, clump size, galactocentric radius, host galaxy luminosity, PSF FWHM, etc. For detection of strong lensing selected objects, at least one more extra variable – lensing magnification is added to the list. We have performed extensive simulations to take all of these effects into account. In our simulations, we assume all clumps will have similar properties to the ones we have detected. If there is a population of clumps at low luminosity that are significantly different in nature than the detected clumps then our completeness corrections derived from our simulations may not as accurately account for the completeness of such different systems. We first fit relations of the rest-frame UV luminosity versus the effective radius for the detected CLASH clumps at 5 redshift bins. We then simulate sets of clumps with assigned absolute AB magnitudes and assign the clump a size that is

randomly drawn from the fitted relations by allowing a scatter of 0.2 dex. To account for effects related to the host galaxies, we use GALFIT to simulate several template host galaxies at different redshift, with luminosity distribution that follows the CLASH observed distribution and place the simulated clumps onto random positions of the template host galaxies. We create the simulated lensed images by re-lensing the combined (galaxy+clump) images to the image plane via CLASH lens models by taking different clump luminosity, redshift and magnification into account. In the image plane, we mimic the observation procedures: first convolve the image with the PSF function of corresponding filter, then bin the lensed images down to the CLASH image resolution (0.065" per pix), then add noise that matches the sky level of CLASH image. We then perform the same image reconstruction procedure on these simulated images in the source plane, as we do to CLASH data. We run the devised clump-finder on the simulated reconstructed images. At each redshift and magnification, the fraction of recovered clumps to the total number is computed as a function of absolute luminosity. We consider 1σ Poisson error as the uncertainty of counting in each luminosity bin.

Figure 4.4 shows the completeness as a function of absolute AB magnitude and magnification factors.

To quantify the detection completeness on stellar mass, we first fit relations of the rest-frame UV luminosity versus the measured stellar mass for the de-

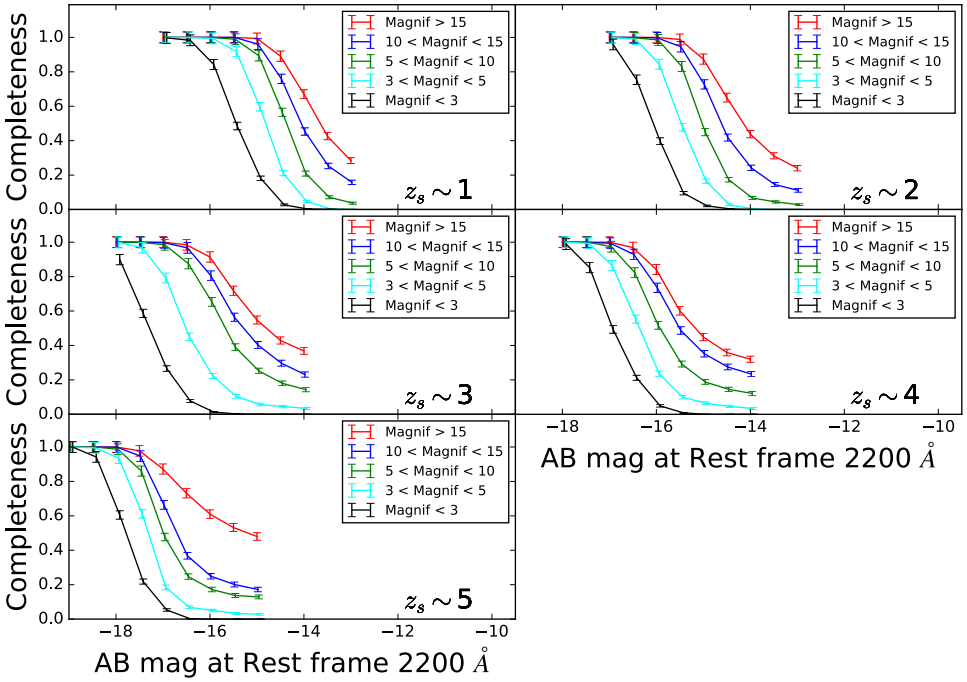


Figure 4.4: The completeness as a function of rest-frame 2200 $\mathring{A}$ absolute magnitude and magnification factor.

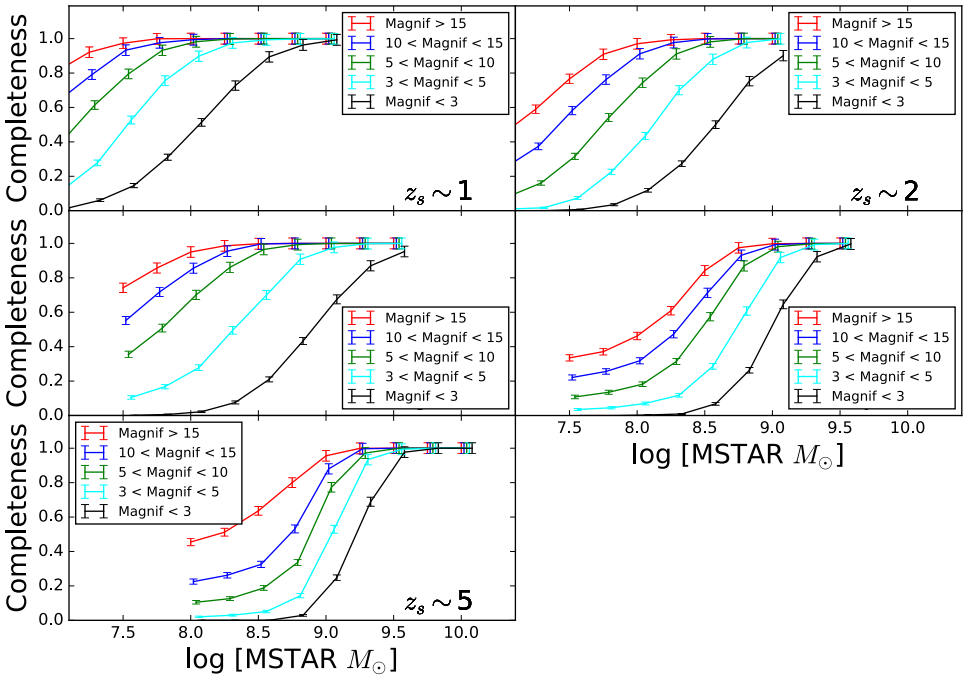


Figure 4.5: The completeness as a function of stellar mass and magnification factor.

tected CLASH clumps at 5 redshift bins. For each assigned stellar mass, an absolute luminosity as well as its size can be randomly drawn from the fitted relations by allowing a scatter of 0.2 dex. Then sets of clumps with assigned absolute luminosity and size that correspond to the assigned mass are simulated. The following simulation procedures and computation of completeness are similar to the described above. Figure 4.5 shows the completeness as a function of stellar mass and magnification factors.

4.6 Physical Properties of Clumps

In this section, we explore the measured physical properties of clumps in detail. The physical properties of clumps are divided into three categories: properties related to total clump mass such as stellar mass and size; properties related to stellar populations such as color, SFR and sSFR, and properties related to compactness such as SFR surface density and stellar mass surface density, and will be discussed separately. These properties will provide us insights about how the clumps assemble their mass via star formation and whether the clumps can survive from the stellar feedback.

4.6.1 Stellar Mass and Size Range of Clumps

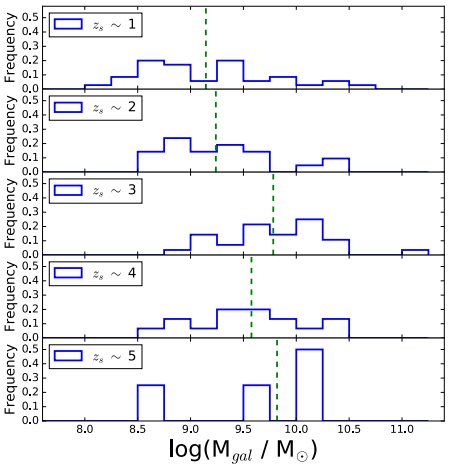
Figure 4.6 lists the distribution of the galaxy and clump stellar mass without incompleteness correction, and clump size distribution with PSF effect corrected at five redshifts. The median mass at $z_s > 3$ is generally larger than that at lower redshifts by 0.5-1 dex. The scatter of the mass distribution at highest redshift is very small ($\sim$ 0.5 dex), which is likely due to the selection effect that only the most massive clumps have been detected. The stellar mass spectrum broadens from $\sim$ 0.5 dex at $z_s \sim 5$ to $\sim$ 3 dex at $z_s \sim 3$, and remains the same level at lower redshifts. Interestingly, a similar increment in median stellar mass with the redshift can also be observed in the stellar mass distribution of

the host galaxies, which implies the evolution in stellar mass of clumps may be correlated with that of the host galaxies.

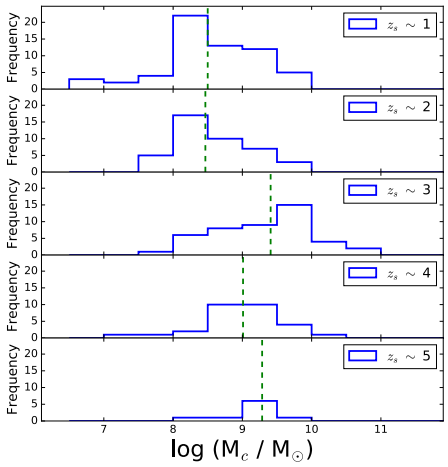
The detected clumps have smallest median size ~ 600 pc at high redshifts ($z_s > 4$). The median size slightly increases to ~ 800 pc as the redshift decreases. The width of the size distribution also increases as the redshift decreases. Therefore, in our sample, the clumps at high redshift appear to be smaller thus more compact; whereas at lower redshifts, the lower mass range the clumps reaches as well as their overall growth in size broaden the explored dynamical range noticeably.

4.6.2 Stellar Population of Clumps

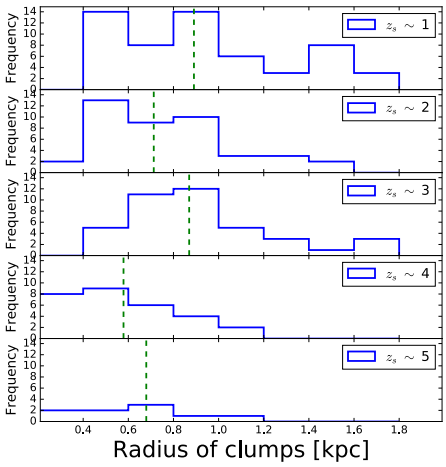
We explore three quantities which related to the stellar population of clumps: rest-frame 2200$\mathring{A}$ - 6500$\mathring{A}$ color, SFR and sSFR. Figure 4.7) lists their distributions in 5 different redshifts. The color and sSFR of clumps remain roughly constant as redshift changes, and the evolution of clump SFR distribution appears to follow that of stellar mass very well. The relatively blue color as well as mild sSFR indicate that the star-forming activity is active over all the redshift ranges, which is not surprising since the selection of the clump is based on rest-frame UV images. The scatter of color, SFR and sSFR distribution increases as redshift decreases, which means that the diversity of the stellar population becomes larger in terms of including more older, star-forming inactive stellar



[a]



[b]



[c]

Figure 4.6: The distributions of galaxy and clump stellar mass, and radius of clumps at 5 redshifts. The green dashed lines are the median values of the distributions.

populations or stellar population with higher metallicity. This increasing scatter in stellar population supports the scenario that the clumps are long-lived, because short lived clumps with age $\sim$ 50 Myr are not expected to have broad scatter and redshift evolution. We will discuss the lifetime of clumps in detail in next section.

We divide the sample of physical properties into 5 redshift bins, and compute the median value and 1σ standard deviation in each redshift bin. Figure 4.8 shows the the redshift evolution of clumps for all the measured physical properties, and the increasing trend in scatter with decreasing redshift can be more clearly inferred from their error bar. It is also interesting to note that the flatness of the sSFR of clumps is reminiscent of the "plateau" of sSFR of star-forming galaxies extending towards high redshifts (Noeske et al., 2007; Daddi et al., 2007; Stark et al., 2009; Gonzalez et al., 2010) (see Figure 4.9), though the uncertainty of the measured sSFR of clumps is very large.

4.6.3 Compactness of Clumps

Previous studies of strongly lensed galaxies have revealed that clumps at high-z appear to be the more massive and compact version of the local HII regions, and their compactness also increases with redshift (Livermore et al., 2012, 2015). Provided the measured physical properties of clumps along with their size, we begin to investigate the surface density of stellar mass and SFR

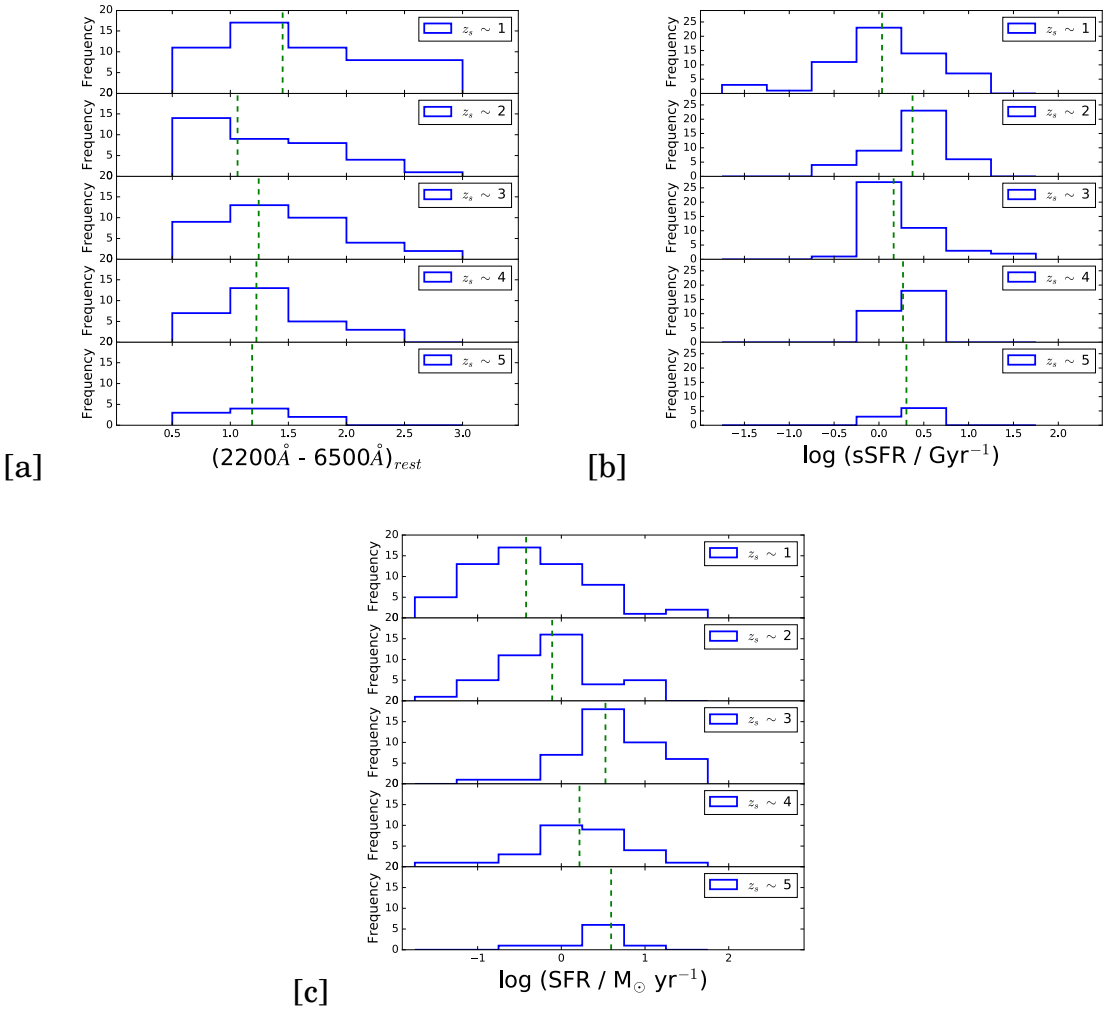


Figure 4.7: The distributions of color, SFR and sSFR of clumps at 5 redshifts. The green dashed lines are the median values of the distributions.

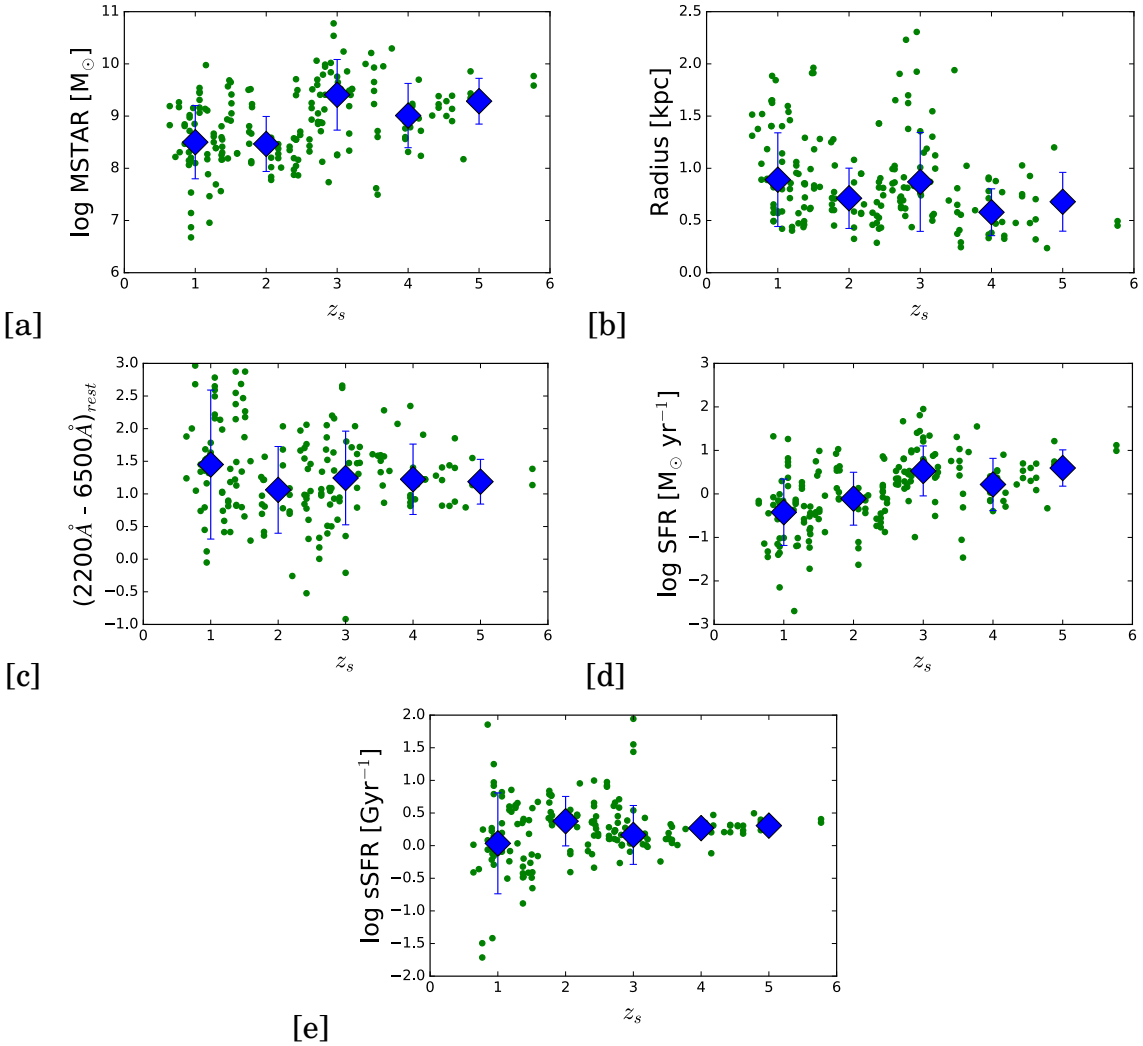


Figure 4.8: The redshift evolutions of the stellar mass, radius, color, SFR and sSFR of clumps. The green dots are the CLASH data. The blue diamonds are the median value within the redshift bin and the error bar is the 1σ standard deviation.

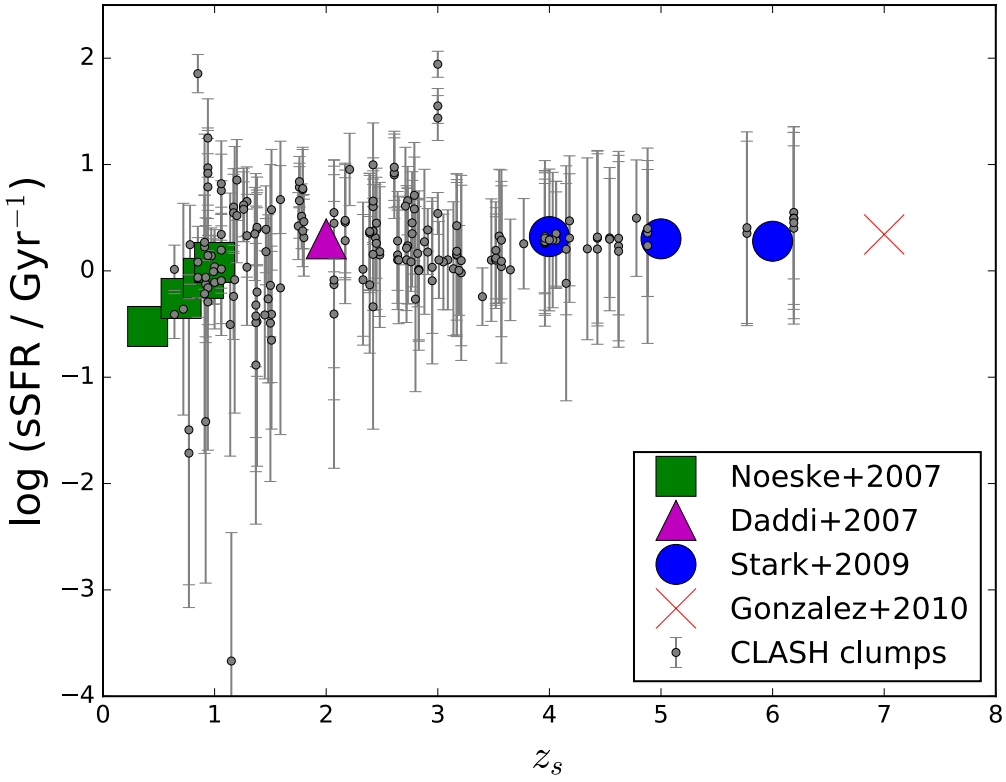


Figure 4.9: Comparison of the evolution of sSFR between clumps and galaxies. Gray points with error bar are clumps from CLASH. Colored markers are measurement of sSFR for high-z galaxies from different studies.

to infer the compactness of clumps as well as their evolution.

We first assign the sample into 5 redshift bins and calculate the median value of surface density of stellar mass and SFR in each bin. We correct the median surface density of stellar mass and SFR for the mass incompleteness and use a parametric form

$$\log\Sigma_{\rm param} = {\rm a} \times \log(1 + {\rm z_s}) + {\rm b} \tag{4.3}$$

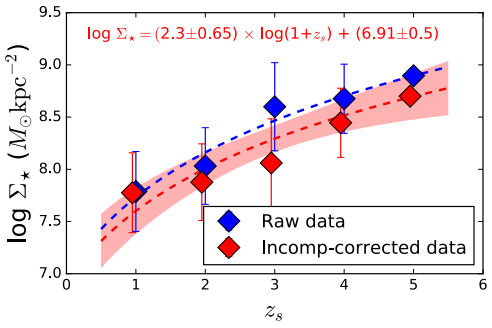
to fit their redshift evolutions. The best fitted lines give:

$$\log\Sigma_{\rm mstar} = (2.36 \pm 0.74) \times \log(1 + {\rm z_s}) + (7.2 \pm 0.51) \tag{4.4}$$

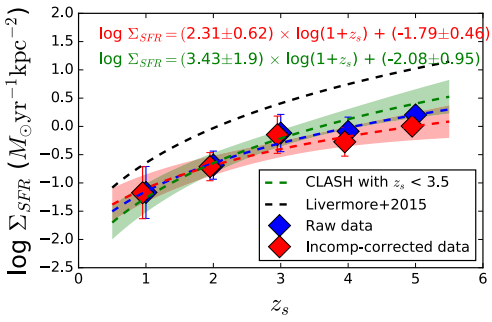
$$\log\Sigma_{\rm sfr} = (2.83 \pm 0.62) \times \log(1 + {\rm z_s}) + (-2.0 \pm 0.46). \tag{4.5}$$

Results are shown in Figure 4.10.

Three features can be immediately observed. First, there are clear increasing trends with redshift for clump surface density of both stellar mass and SFR. Second, the clump stellar mass surface density at low redshift ($z < 2$) is comparable to that of local GMC $\sim 10^8 {\rm M_\odot kpc^{-2}}$ (Larson, 1981) while the clump SFR surface density is higher by $\sim$ 0.5 dex. Third, the clump surface density of stellar mass and SFR at $z_s > 3$ are typical 5-10 times higher than those at $z_s \sim 1$. These results confirms the previous results that the clumps at higher redshift are not only more massive, but also more compact and star-forming



[a]



[b]

Figure 4.10: The redshift evolutions of surface density of stellar mass and SFR of CLASH clumps. The blue and red dashed curves in (a) and (b) are the best fitted lines for the raw data and incompleteness-corrected data; the green dashed curve in (b) is the best fitted line for $z_s < 3$ only; the black dashed curve is the best fitted line from Livermore et al. (2015). The blue and red diamonds in (a) and (b) are raw and incompleteness-corrected CLASH data and the error bar denotes the 1σ standard deviation. The color shaded regions are the 1σ scatter of the best fitted lines.

active than those at lower redshift.

We compare our results with Livermore et al. (2015), who studied a sample of 50 clumps identified from 17 star-forming galaxies spanning redshift range $1.28 < z_s < 3.5$. Their selection criteria on clump detection is 3σ above the diffuse galaxy background level, which is similar as ours. We first compare the size and SFR distributions of clumps between the samples. As Figure 4.11 lists, their sample preferentially selects smaller clumps than ours, thus with smaller SFR. e.g. their median radius and SFR are 2.5 and 3 times smaller ours, respectively. There could be three potential sources of bias: 1. different methods to measure the clump size: they measure the clump size based on an isophotal method whereas we use a core method to infer the size; 2. different methods

to correcting the PSF blurring: we correct the blurring effects based on results from simulations, while they directly subtract the area of effective source plane PSF in quadrature; 3. most importantly, their sample of host galaxies are selected to be highly lensed in the image plane, thus potentially bias towards high magnification factor. e. g. the median magnification of their host galaxies is ~ 24, a factor of 5 larger than our median value. Bear in mind that a higher overall magnification factor would lead to a higher average S/N ratio in the image plane and finer average spatial resolution in the source plane, which allows more detections of smaller clumps. To test this, we divide the CLASH clumps into 2 subsamples: clumps from host galaxy with magnification factor $>= 9$ and < 9, and compare their size distributions. From Figure 4.12, we indeed observe that the smaller clumps are preferentially detected from galaxies with higher magnification. Therefore, the bias in size measurement between two studies is more likely due to the selection bias on magnification of the host galaxies.

Next, we compare our measured $\Sigma_{SFR} - z_s$ relation with the result from Livermore et al. (2015), which gives:

$$\log\Sigma_{\rm SFR} = (3.5 \pm 0.5) \times \log(1 + {\rm z_s}) + (-1.7 \pm 0.2). \quad (4.6)$$

The slope and intercept are consistent with ours, within the uncertainties. We

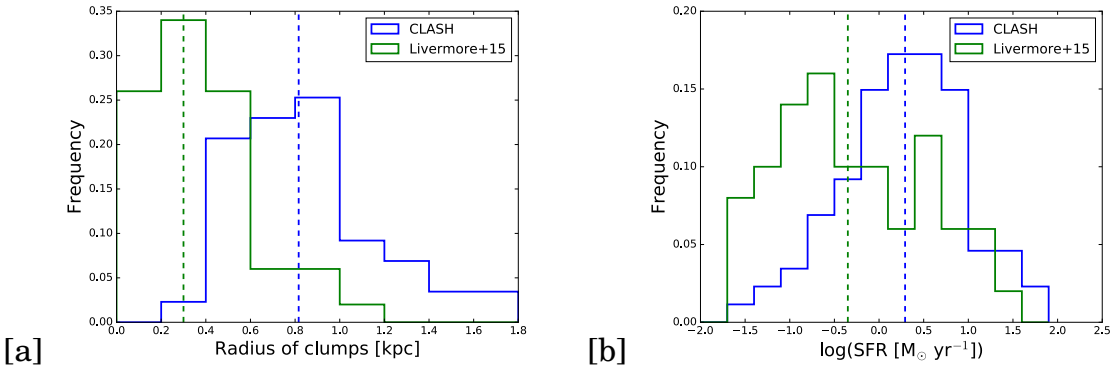


Figure 4.11: The comparison between the distributions of the measured half-light radius and SFR in this study and Livermore et al. (2015). The blue and green dashed lines denote the median value of the distributions.

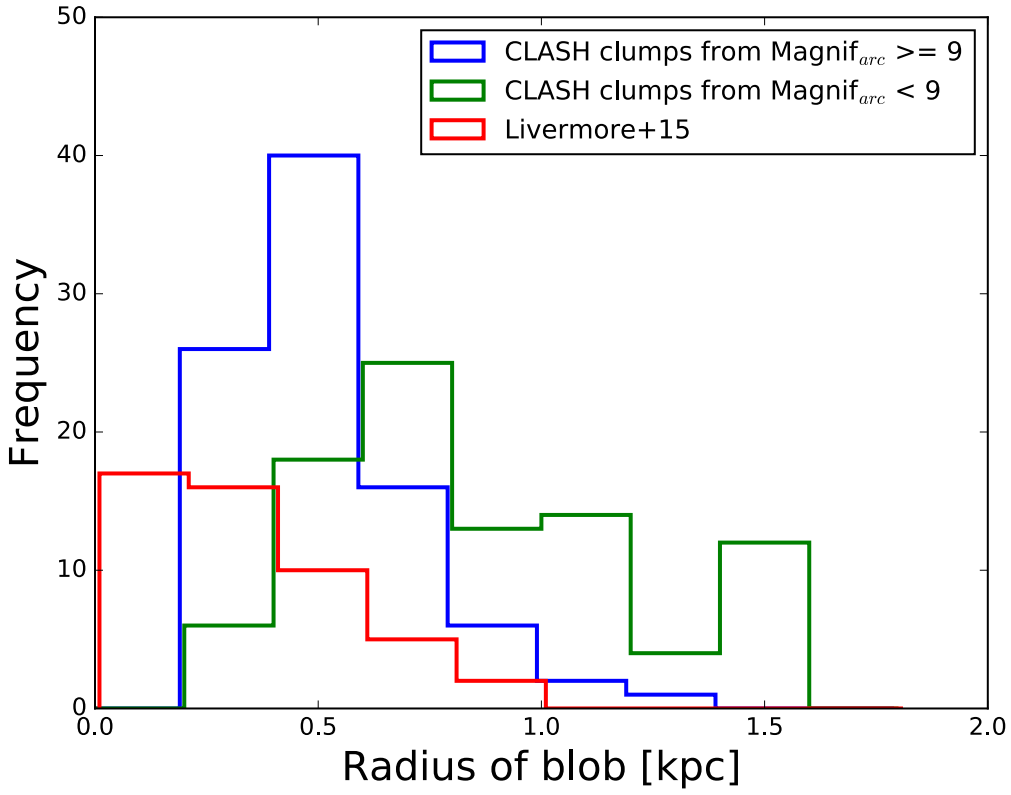


Figure 4.12: The comparison between the distributions of the measured half-light radius of CLASH clumps with different magnification and clumps from Livermore et al. (2015).

note that their sample only includes 4 star-forming galaxies at $z_s > 3$, and only one with $z_s > 3.5$. For consistency, we exclude our data with $z_s >= 3.5$ and re-fit the relation. As Figure 4.12 shows, the best re-fiited line for CLASH sample gives:

$$\log\Sigma_{\rm SFR} = (3.43 \pm 1.9) \times \log(1 + {\rm z_s}) + (-2.08 \pm 0.95). \quad (4.7)$$

The re-fitted result remains consistent with their result, though with larger uncertainties. Therefore, the evolution of compactness of the clumps appears to be less sensitive to the selection of clumps on different scales.

4.7 Fundamental Planes of Clumps

A fundamental plane is a set of bivariate correlations connecting some of the physical properties of normal elliptical or spheroidal galaxies, which include their luminosity,mass, radius, velocity dispersion, surface brightness, color, density, etc. The position on the fundamental plane reflects the dynamical properties of the stellar systems, which is extremely useful to distinguish the stellar systems with different dynamical properties but similar morphology as exemplified, for instance, by elliptical galaxies and dwarf spheroidals (Kormendy et al., 2009). Therefore, to see whether the clumps which have spheroidal morphology share similar dynamical properties with other stellar systems with similar morphology, we study the fundamental plane of the clumps

and compare their positions with other stellar systems with spheroidal morphology. The fundamental plane is usually expressed as relationships between the effective radius, average surface brightness and central velocity dispersion of normal elliptical galaxies. In this study, we choose to use a different combinations: stellar mass, surface density of stellar mass and effective radius, because we do not have kinematic information on the clumps and thus cannot include a velocity dispersion measurements in the current study.

4.7.1 Samples Description

The sample of stellar systems for comparison is mainly taken from Table 1 in Misgeld & Hilker (2011), which lists a wide range of spheroidal stellar systems from old elliptical galaxies and bulge, dwarf spheroidals, compact ellipticals, globular clusters, star clusters, ultra compact dwarfs and local dwarf galaxies. The details in stellar mass estimation of different stellar systems can be found in section 2.1 in Misgeld & Hilker (2011). We overplot the CLASH high-z clumps along with high-z clumps detected from other studies (Wisnioski et al., 2012; Swinbank et al., 2009, 2012; Livermore et al., 2012, 2015) and high-z globular clusters (Rigby et al., 2017; Vanzella et al., 2017). Since in most of studies of clumps at high-z, only values of SFR are provided, we convert the SFR values in the literature to stellar mass by adopting a very conservative sSFR $\sim 3\mathrm{Gyr}^{-1}$ which is the median value of CLASH sample. Most of the high-

z clumps in these studies are star-forming clumps, the actual value of sSFR could be higher than the value we use (Guo et al., 2012; Wuyts et al., 2014), which thus leads to a smaller stellar mass and surface mass density. Therefore, the estimated stellar mass based on sSFR is only an upper bound for those high-z clumps without measured stellar mass.

4.7.2 Results

Figure 4.13, 4.14 and 4.15 show the fundamental planes of local and high-z spheroidal stellar systems. It can be clearly seen that stellar systems can be categorized into two big branches: galaxies or components of galaxy including the dwarf galaxies, dwarf spheroidals, bulges and elliptical galaxies; and smaller stellar complex including the star clusters, globular clusters and ultra compact dwarfs. Surprisingly, all the high-z clumps appears to fall onto the branch of galaxies, which is at odds with the prior impression that clumps are bigger version of local HII regions which are possibly transient features as well. The high-z clumps overlap very well with the dwarf spheroidals and late type bulges on the fundamental planes. In specific, the CLASH clumps as well as the clumps from Wisnioski et al. (2012) and Swinbank et al. (2012) mainly overlap with more massive dwarf spheroidals whereas the clumps from Livermore et al. (2012, 2015) are consistent with less massive dwarf spheroidals. The smaller clumps detected as high-z globular clusters by Rigby et al. (2017)

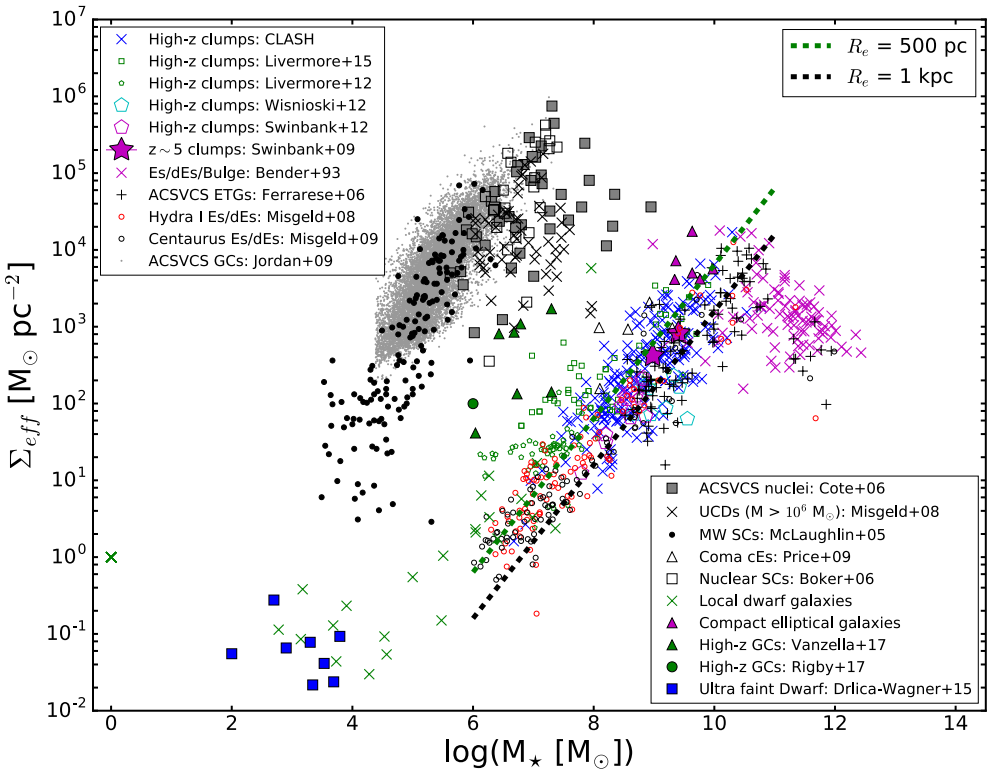


Figure 4.13: The surface mass density versus stellar mass for all the spheroidal stellar systems in the sample. The green and black dashed lines denote a constant effective radius of R_e = 500 pc and 1 kpc for reference.

and Vanzella et al. (2017) fall in between the two branches, and some of them overlap lie at similar position as the ultra compact dwarfs identified in galaxy clusters. If a higher sSFR is assumed, these points would shift towards left and bottom, thus well coincide with local star clusters on the fundamental planes.

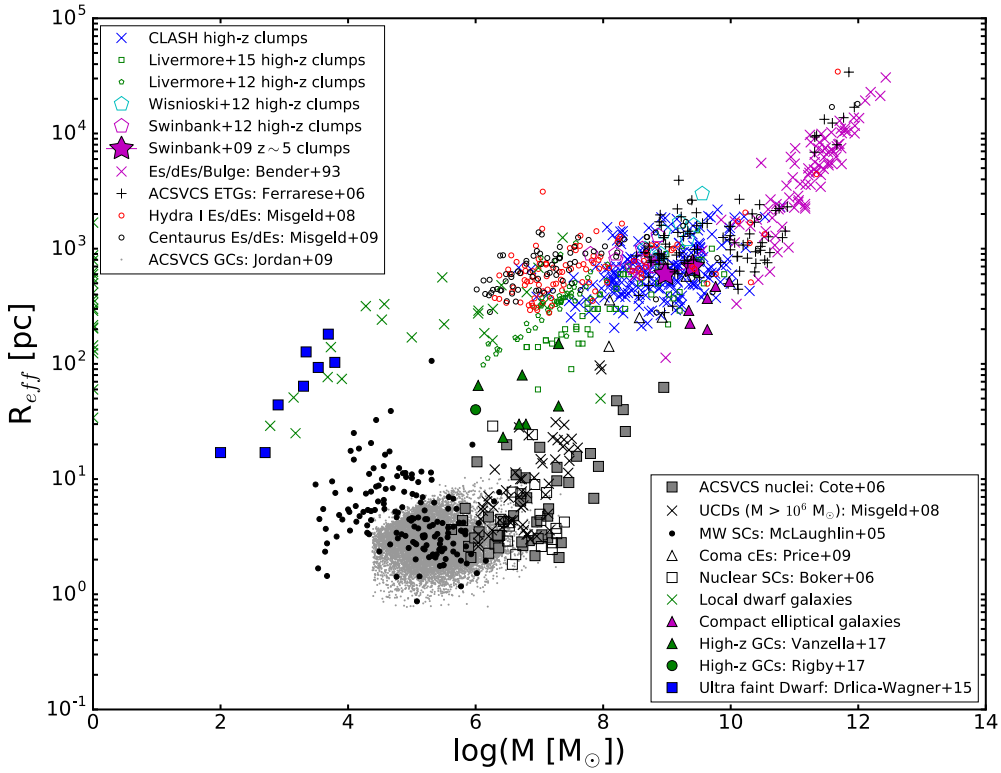


Figure 4.14: The effective radius versus stellar mass for all the spheroidal stellar systems in the sample.

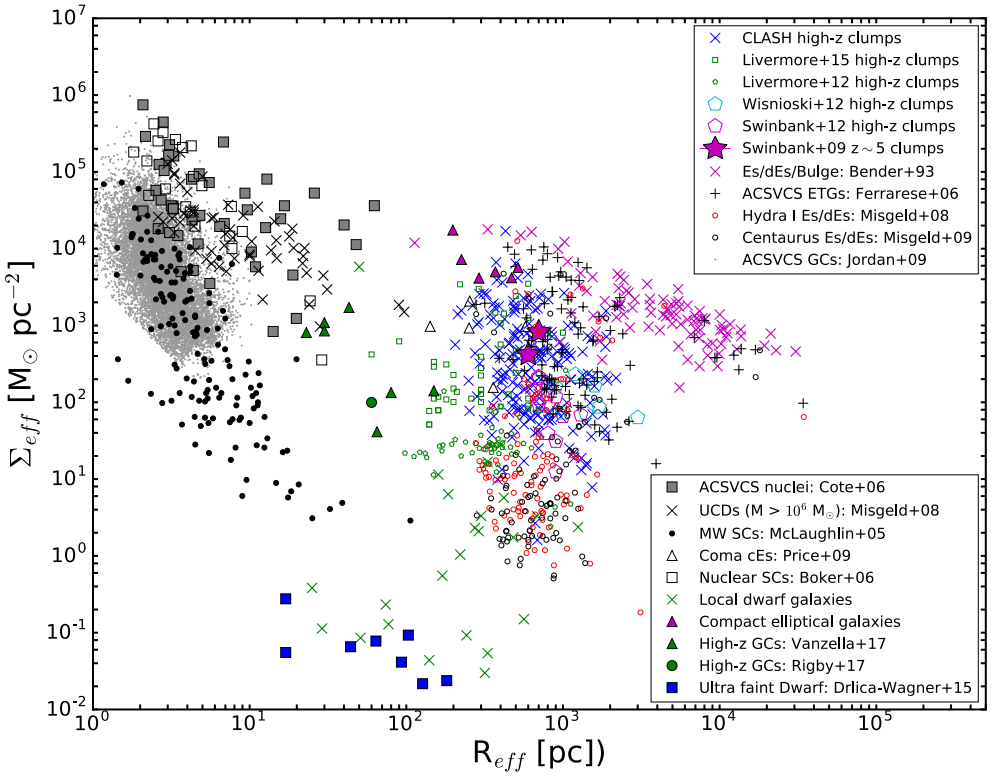


Figure 4.15: The surface mass density versus effective radius for all the spheroidal stellar systems in the sample.

4.7.3 Implications for Galaxy Evolution

It has been hypothesized for decades that the environmental effects might play a crucial role in the origin of the dwarf spheroidals (Kormendy, 1985; Binggeli et al., 1990; Beasley et al., 2006; Smith et al., 2009; Bialas et al., 2015), in the sense that the dwarf spheroidals are the disk galaxies with gas removed efficiently and star-forming ceased by certain environmental effects. There are several evidence to support this hypothesis: first, the morphology of the galaxy and the fractional mass in spheroidal component in local group is a strong function of the projected distance to the group center (Moffett et al., 2016). Most of the irregular dwarf galaxies are found to be in the outskirt of the group, while the early type galaxies preferentially locate at small projected distance; second, the structural studies on the late type dwarf galaxies have shown that the size of the late type dwarfs is noticeably smaller in high local density than in low local density, which lie closer in size to those of the early type dwarfs (Janz et al., 2016); third, residual star formation activity with much flatter shape has been found in dwarf spheroidals outside the densest environments (Lisker et al., 2006; Tully & Trentham, 2008). Current efforts on studying the origin of the dwarf spheroidals mainly focus on evolutionarily linking the dwarf spheroidals and dwarf irregulars. In fact, Elmegreen et al. (2009) found a striking resemblance of the morphology of high-z clumpy galaxies and local dwarf irregualrs as well as their similar ratio of gas velocity dispersion to the circu-

lar velocity σ_{gas}/V_{circ}. Therefore, it is well possible that similar environment effects that act on the high-z clumpy galaxies would result in similar morphological transformation. On the other hand, the dynamical properties of clumps are similar to those of dwarf spheroidals as seen from the fundamental planes. The main differences between the two stellar systems are their stellar population and environment: the high-z clumps are typically star-forming active system with high sSFR that are embedded in irregular galaxies, which are commonly found in the field; while the dwarf spheroidals are typically quenched, red galaxies, which are preferentially found in galaxy clusters or groups as satellite galaxies. Therefore, if there exists potential relations or evolutionary track between the high-z clumps and local dwarf spheroidals, the cessation of the star-forming activity in clumps should be closely related to their environments. There are indeed various proposed environmental quenching mechanisms such as ram-pressure stripping (Dressler & Gunn, 1983), strangulation (Larson et al., 1980) and tidal harassment (Moore et al., 1996), that are responsible for the quenching of star formation in satellite galaxies by removing their gas efficiently. Therefore, similar to the proposed morphological transformations between the dwarf spheroidals and dwarf irregulars, there also likely exists an interesting evolutionary track that: when the clumpy irregular galaxies fall into some dense environments such as galaxy clusters or groups, some environmental effects such as ram-pressure stripping would efficiently remove

the gas from the infalling galaxies. Thus only the most compact components of the progenitor infalling galaxies that have been observed as clumps at high-z, survive to the environmental effects. The removal of the gas then quenches the star formation in remnant clumps, thus observed as red dwarf spheroidals. Note that a necessity of this speculation is that the observed high-z clumps need to be sufficiently compact and long-lived to survive from external environmental transformation and internal stellar feedback. We will discuss the longevity of the clump in the next Chapter in detail.

Moreover, Barro et al. (2017) study the evolution of the scaling relation between the effective density (Σ_e, $r < r_e$) and core density (Σ_1, $r < 1$ kpc) to the stellar mass of a sample of star-forming galaxies from CANDELS. They find that the $\Sigma_1 - \mathrm{M}_\star$ relation is tighter than $\Sigma_\mathrm{e} - \mathrm{M}_\star$ relation. Both relations appear to exhibit almost constant slope and scatter since $z_s \sim 3$ and the scatter of the relation is much smaller than the mass range the sample spans. They suggest that there could be an evolutionary track for SFGs that these SFGs assemble their mass along these scaling relations. On the other hand, we plot reference lines of constant effective radius of $\mathrm{R_e}$ = 500 pc and 1kpc in Figure 4.13. It appears that most of the high-z clumps and dwarf spheroidals follow the $\mathrm{R_e}$ = 500 pc tightly. The dispersion to the line for clumps is similar to that for dwarf spheroidals, which is ~ 0.25 dex. The clumps combined with the dwarf spheroidals span 4 dex in stellar mass which is much larger than the

dispersion. We propose that there could be an approximate evolutionary track along the constant effective radius line for clumps as well. The implication of this evolutionary track is that the stellar systems such as progenitors of dwarf spheroidals could build up their mass budget in a smooth and regular way as long as they continue to form stars, without significant growth in size. This is also consistent with the inside-out growth picture of the galactic disk due to the accretion of material with higher angular momentum and star formation which lead to the steady increase in Σ_c along with the galaxy stellar mass. To confirm that, we compare the core density[2] and stellar mass of CLASH lensed galaxies with the scaling relations from Barro et al. (2017) in Figure 4.16. The best-fit scaling relation for CLASH lensed galaxies at $0.5 < z_s < 3.5$ gives:

$$\log(\Sigma_{\rm c} \ / \ {\rm M}_\odot \ {\rm pc}^{-2}) = (0.727 \pm 0.126) \times \log({\rm M}_{\rm gal} \ / \ {\rm M}_\odot) + (3.079 \pm 0.157), \quad (4.8)$$

which is well consistent with the scaling relation from Barro et al. (2017) at similar redshift ranges. In addition, we overplot the predicted evolutionary track at $z_s = 2$ (the gray belt) based on Vela set of 35 hydrodynamical simulations described in Zolotov et al. (2015) (also in Ceverino et al. (2015)), which also matches our scaling relation well. The consistency between our relation with those from other observations and simulations in predicting the evolu-

[2]We adopt the surface mass density of the most massive clump in each galaxy as their core density. For galaxy with the effective radius of the most massive clump less than 1 kpc, this assumption may over-estimate their core density.

tionary track for SFGs or the host galaxies of clumps lends us confidence that an similar evolutionary track could also apply to clumps.

4.8 Rest-frame UV Luminosity Function of Clumps

Distribution functions are broadly used quantities to enable a direct comparison between the theory and the observations. The distribution function of galaxy rest-frame UV luminosity is in particular useful and straightforward means to explore the star-forming activity in the local and distant Universe, which has been studied over a wide range of redshifts (Bouwens et al., 2015; Bowler et al., 2015; Finkelstein et al., 2015; Livermore et al., 2017), thus provides us the detailed insight into the physical processes in galaxy evolution and the reionization of the early Universe. On the other hand, as we trace back to the early stage of the Universe, the morphology of the galaxies become more clumpy and the sub-galactic structures appears to stand out more from the overall flux. To understand the formation mechanism and evolution of these sub-galactic structures, it is natural to step further to investigate the rest-frame UV luminosity function of the clumps, as the key ingredient of the sub-galactic structure. In this section, we will describe the procedure to evaluate the rest-frame UV luminosity function of clumps.

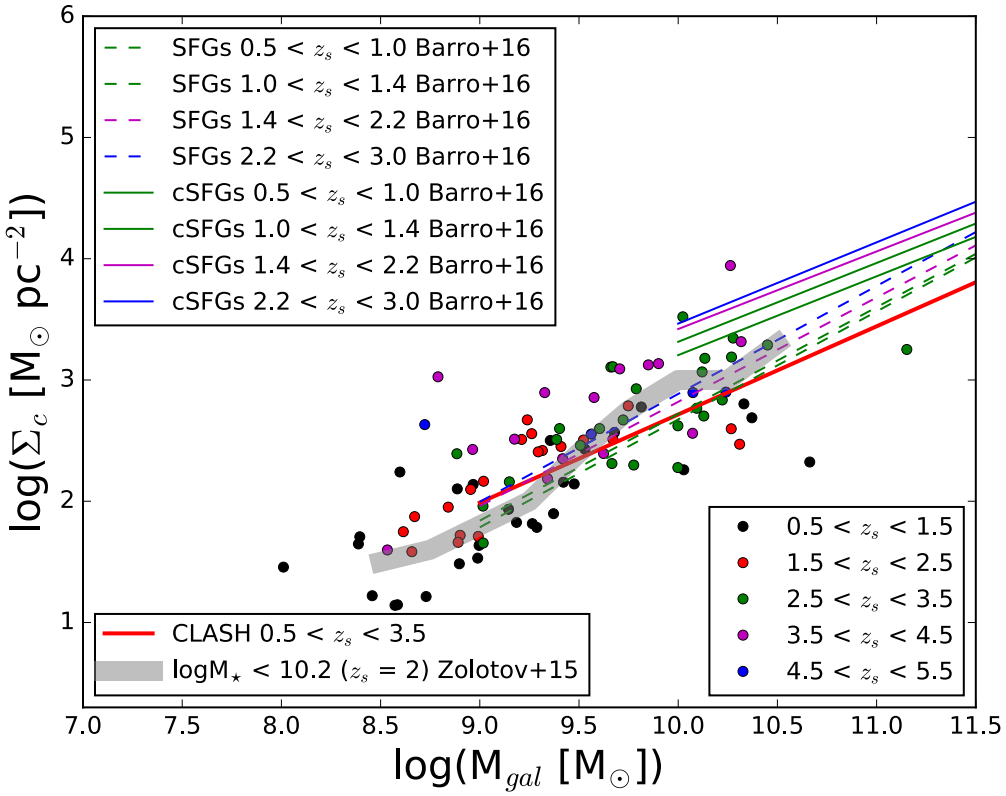


Figure 4.16: Comparison of core surface mass density Σ_c and galaxy stellar mass $M_\star$ between CLASH lensed sample, CANDELS sample from Barro et al. (2017) and the predicted evolutionary track from Zolotov et al. (2015). The colored dots are the CLASH data with galaxy stellar mass and the surface mass density of the most massive clump in that galaxy at 5 redshifts. The red solid line is the best-fit line for CLASH data at $0.5 < z_s < 3.5$. The colored dashed and solid lines are the best-fit scaling relation for SFGs and compact SFGs at four redshifts from Barro et al. (2017). The gray belt marks the predicted evolutionary track for galaxies with mass less than $10^{10.2}M_\odot$ at $z_s = 2$ from Zolotov et al. (2015).

4.8.1 Incompleteness Correction

We divide the whole clump sample into 5 redshift bins. In each redshift bin, we assign the data into 10 absolute magnitude bins with bin width set to 0.5 mag. To account for the magnification, we further assign the data in each magnitude bin into 4 magnification bins: $A_c <= 1.5$, $1.5 < A_c <= 2.5$, $2.5 < A_c <= 3.5$ and $3.5 < A_c <= 4.5$, where A_c is the linear magnification of clumps. In each magnitude bin, we correct the magnification dependent incompleteness for each data and sum the counts up as the corrected count in the corresponding bin. Following this, we correct the incompleteness at each magnitude and redshift bin.

4.8.2 Uncertainties

To obtain reliable accuracy and robustness of the luminosity function, there are several sources of uncertainties to be taken into account: first, 1σ Poisson error $\sigma_{Poisson}$ is computed for each magnitude bin; second, we estimate the uncertainty contribution σ_{SED} from the photometric redshift estimation and the SED fitting by conducting 200 Monte Carlo simulations on the catalog. For each realization, we perturb the photometric redshift and rest-frame 2200$\mathring{A}$ AB magnitude within their 1σ scatter value estimated from BPZ and iSED-fit, and recalculated the luminosity function. The 1σ scatter in the resulting

luminosity function is then adopted as the redshift and absolute luminosity dependent uncertainty; 3. we also estimate the uncertainty from the cosmic variance σ_{cv}. We use the redshift and mass dependent cosmic variance catalog provided in Moster et al. (2013). Since the catalog in Moster et al. (2013) relates the estimated cosmic variance to the stellar mass of galaxies not rest-frame UV luminosity, we compute the median value of the stellar mass of host galaxies in each UV magnitude bin, and find the corresponding cosmic variance based on this median mass from the Moster's catalog. The total uncertainty is then:

$$\sigma_{\rm tot} = \sqrt{\sigma^2_{\rm Poisson} + \sigma^2_{\rm SED} + \sigma^2_{\rm cv}} \tag{4.9}$$

4.8.3 Parameterizations of Fitting

Observations have shown that the galaxy luminosity function follows a typical characteristic shape with an exponential decline at the bright end and a power-law slope at the faint end, transitioning at a characteristic luminosity. Such shape can be well parameterized by function with three parameters as proposed by (Schechter, 1976): the characteristic luminosity or magnitude ($L_\star$ or $M_\star$), the power-law slope α and the normalized overall number density of galaxies $\phi_\star$. Motivated by the possibility that the mass distribution of clumps follows that of the host galaxies (Section 3.6.1), we adopt the Schechter parameterization which is prevalent in fitting the galaxy luminosity function, to fit

the rest-frame 2200$\mathring{A}$ luminosity function of clumps. The parameterizations in terms of magnitude are given in Equation 4.10:

$$\phi(M) = 0.4\ln(10)\phi_\star 10^{-0.4(\mathrm{M}-\mathrm{M}_\star)(\alpha+1)}\mathrm{e}^{-10^{-0.4(\mathrm{M}-\mathrm{M}_\star)}} \tag{4.10}$$

4.8.4 Results

We fit the completeness corrected rest-frame 2200 $\mathring{A}$ luminosity function of clumps at 5 redshift bins, down to AB M_{2200A} = -15, -15, -15.5, -15.5, -16 for $z_s \sim 1, 2, 3, 4, 5$. The results are shown in Figure 4.17, and the best-fit Schechter parameters as well as the goodness of the fit are summarized in Table 4.2.

Table 4.2: Best-fit Schechter Parameters for Rest-frame UV Luminosity Function. $\Phi_c^\star$ is in units of Mpc^{-3} mag^{-1}.

Redshift	Log($\Phi_c^\star$)	α	log($\mathrm{M}_c^\star$)	χ^2_ν
$0.5 < z_s < 1.5$	-2.14 $\pm$ 0.16	-0.2 $\pm$ 1.11	-16.12 $\pm$ 0.76	2.57
$1.5 < z_s < 2.5$	-2.58 $\pm$ 0.14	-0.61 $\pm$ 0.36	-18.27 $\pm$ 0.67	1.15
$2.5 < z_s < 3.5$	-2.60 $\pm$ 0.19	-0.78 $\pm$ 0.3	-19.98 $\pm$ 0.82	1.22
$3.5 < z_s < 4.5$	-2.82 $\pm$ 0.32	-0.86 $\pm$ 0.48	-18.92 $\pm$ 1.11	1.30
$4.5 < z_s < 5.5$	-3.03 $\pm$ 0.63	-1.09 $\pm$ 0.71	-17.98 $\pm$ 2.25	1.82

The reduced χ^2 values indicate that a single Schechter function is sufficient to describe the shape of the distribution. The trend of the best-fit Schechter function can be clearly seen from Figure 4.18: the normalized number density $\Phi_c^\star$ increases as the redshift decreases. The characteristic magnitude M_{2200A} reaches its maximum value at $z_s \sim 3$ and deceases significantly towards lower

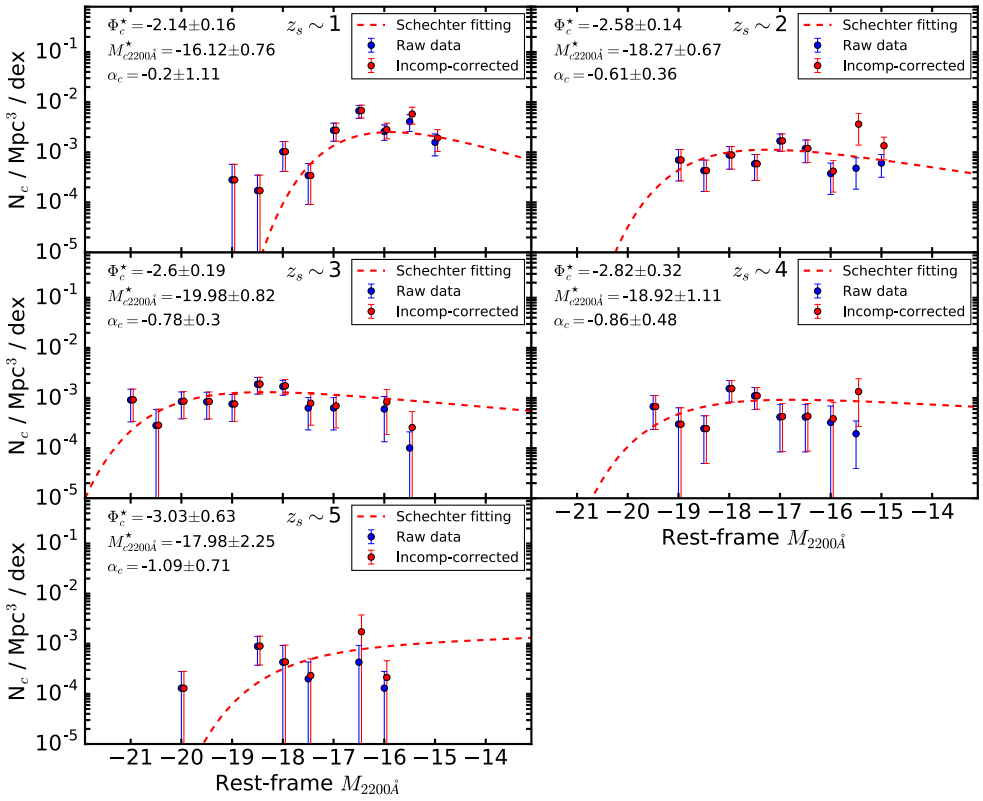


Figure 4.17: The incompleteness corrected rest-frame 2200 $\AA$ luminosity function of clumps at 5 redshift bins. The blue points are raw data and red points are the incompleteness corrected data. The red dashed lines are the best-fitted function by Schechter parameterization.

redshift, while it also slightly decreases towards the highest redshift. The trend is similar to what we observed in evolution of clump stellar mass in Section 3.6.1, which is consistent with the picture that the evolution of star formation in clumps is related to their host galaxies. There is only weak increase in α as redshift decreases, given the large error bar, but this trend is also consistent with that of the galaxies: the faint-end slope becomes steeper as the redshift increases (Bouwens et al., 2015; Finkelstein et al., 2015; Bowler et al., 2015). We note that the faint end slope for clumps is systematically shallower than that of the galaxies, which typically has the value of $\sim$ -1.6 – -2.0 at similar redshifts. It could suggest that the faint clumps are less dominated in galactic UV flux budget than the faint galaxies does in the cosmic UV flux budget in the entire baryonic hierarchy because there is no physical motivation to suggest these two should converge. However, since the slope is most affected by the faint end data, the insufficient data in the lowest magnitude bins which potentially leads to incorrect incompleteness correction would also bias the estimation of the faint end slope. Therefore, future larger clump samples with lower limiting magnitude are needed to assess this problem.

For convenience, we sketch the Schechter functions with best-fit parameters at different redshifts in Figure 4.19.

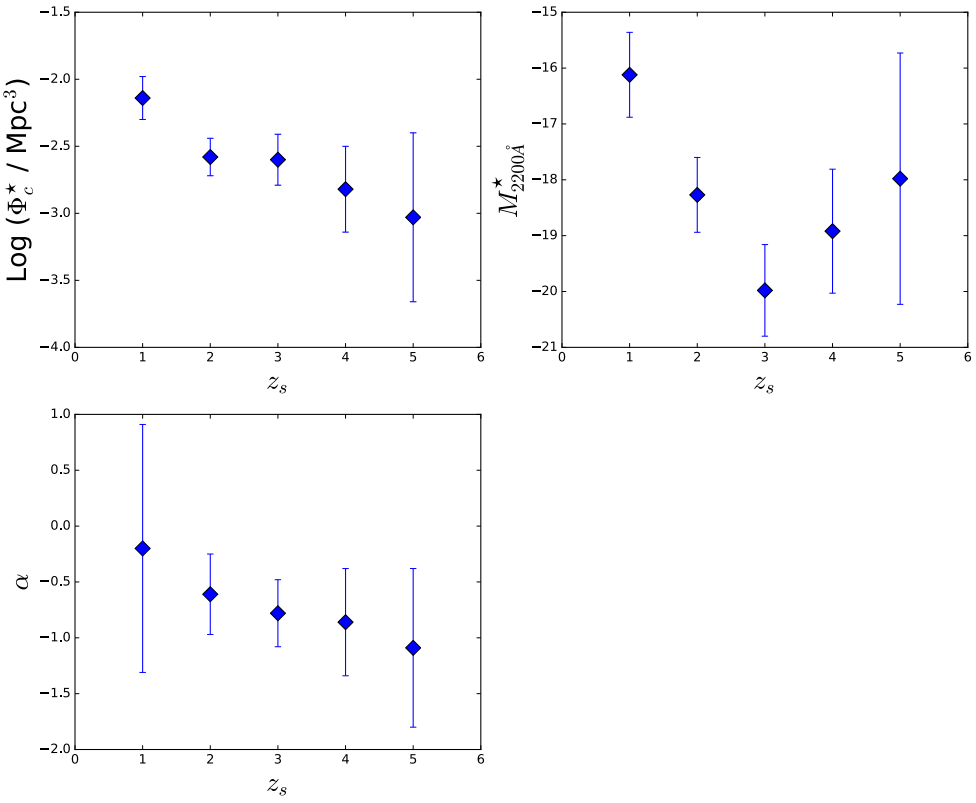


Figure 4.18: The redshift evolution of the best-fit Schechter parameter of rest-frame 2200$\mathring{A}$ luminosity function. The error bar denotes 1σ rms error.

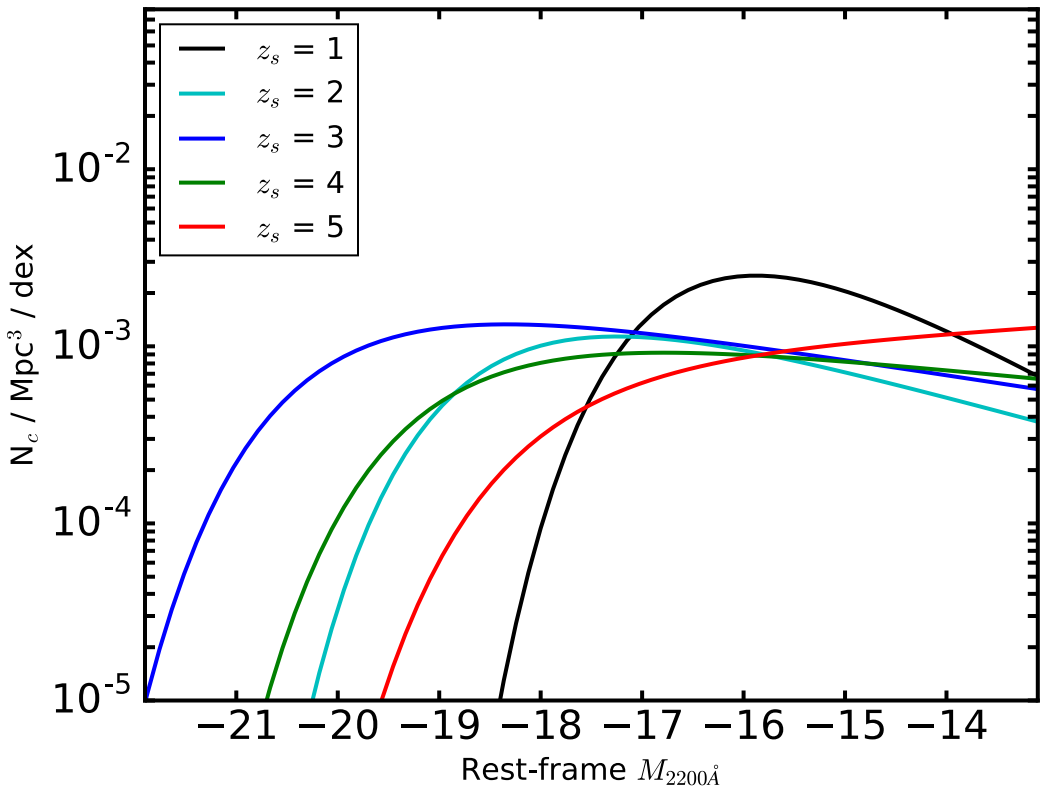


Figure 4.19: The rest-frame $2200\mathring{A}$ luminosity functions with best-fit Schetchter parameters at 5 different redshifts.

4.9 Stellar Mass Function of Clumps

Stellar mass function and its evolution are crucial to study the history of the mass assembly of the stellar systems. It is in particular important for study of clumps because the clumps are proposed to play key role in the assembly of the galactic bulges, and a direct comparison between the stellar mass functions of the two would be necessary to assess the feasibility of the scenario. To our best knowledge, a direct measurement on stellar mass function of clumps is still lacking. Therefore, in this section, we describe the methodology to compute the stellar mass function of clumps. For stellar mass function of clumps, we still adopt a Schechter parameterization:

$$\phi(M) = \phi_{\star} 10^{(\mathrm{M}-\mathrm{M}_{\star})(\alpha+1)} \mathrm{e}^{-10^{(\mathrm{M}-\mathrm{M}_{\star})}}, \tag{4.11}$$

where $\phi_{\star}$ is the normalized number density, $\mathrm{M}_{\star}$ is the characteristic stellar mass and α is the low mass end slope.

4.9.1 Incompleteness Correction

We assign the data into 10 stellar mass bins with bin width = 0.25 dex in each of 5 redshift bins. Similar to that in computing the luminosity function, we correct the data in 4 linear magnification bins in each stellar mass bin to

account for the lensing magnification. The count in 4 magnification bins are then summed up as a total corrected count in each stellar mass bin.

4.9.2 Uncertainties

Similar to the procedure in computing the luminosity function. We account for uncertainties from three source in the estimation of stellar mass function: the statistical uncertainty (Poisson error) $\sigma_{Poisson}$; the uncertainty from the estimation of the photometric redshift and SED modeling σ_{SED}; the uncertainty from the cosmic variance σ_{cv}. The estimation of σ_{SED} is based on conducting Monte Carlo simulations, which is described in Section 3.8.2. We calculate the median stellar mass of the host galaxy of clumps in each stellar mass and redshift bin, and use the mass and redshift dependent fraction error from cosmic variance based on the catalog from Moster et al. (2013). The total uncertainty is given in Equation 4.9.

4.9.3 Results

Figure 4.20 lists the incompleteness corrected stellar mass function of clumps at 5 redshift bins. The best-fit Schechter parameters are included in Table 4.3. The reduced χ^2 values show that the single Schechter function only gives good fit at $z_s \sim 4, 5$. The fitting turns worse mainly because of the deficits around

certain mass at $z_s \sim 1, 2$ and upturn at low mass end at $z_s \sim 3$, which is likely the statistical fluke or insufficient sampling of host galaxies at certain mass range. The evolution of the Schechter parameters are sketched in Figure 4.21. Except for $z_s \sim 1$, the trend in number density of clumps is consistent with no evolution, while at $z_s \sim 1$, the number density becomes larger. The trend in characteristic stellar mass is similar to that in characteristic luminosity: it peaks at $z_s \sim 3$ and decreases towards low and high redshifts, which also appears to follow the stellar mass distribution of the host galaxies. Interestingly, the trend in low mass end slope at $z_s > 2$ is broadly consistent with flat ($\alpha \sim -1$), which implies the clumps appear to form in a scale independent way at high redshifts.

Table 4.3: Best-fit Schechter Parameters of Stellar Mass Function. $\Phi_c^\star$ is in unit of Mpc^{-3} dex^{-1}, and $M_c^\star$ is in unit of M$_\odot$.

Redshift	Log($\Phi_c^\star$)	α	log($M_c^\star$)	χ^2_ν
$0.5 < z_s < 1.5$	-2.13 $\pm$ 0.09	-0.36 $\pm$ 0.33	9.03 $\pm$ 0.15	5.54
$1.5 < z_s < 2.5$	-2.82 $\pm$ 0.24	-0.9 $\pm$ 0.26	9.65 $\pm$ 0.29	3.71
$2.5 < z_s < 3.5$	-2.75 $\pm$ 0.27	-0.87 $\pm$ 0.31	10.0 $\pm$ 0.43	4.28
$3.5 < z_s < 4.5$	-3.01 $\pm$ 0.36	-1.11 $\pm$ 0.31	10.02 $\pm$ 0.44	1.41
$4.5 < z_s < 5.5$	-3.04 $\pm$ 0.5	-1.12 $\pm$ 0.7	9.49 $\pm$ 0.65	0.78

We show the stellar mass functions of clumps with best-fit Schechter parameters at 2 redshifts $z_s \sim 3, 4$ in Figure 4.22. We also plot the local stellar mass function for the bulges in early and intermediate/late type spiral galaxies and S0 galaxies (Moffett et al., 2016) for comparison. The median mass of host galaxies of massive clumps at $z_s \sim 3, 4$ is about $10^{10.2} M_\odot$, which is likely the

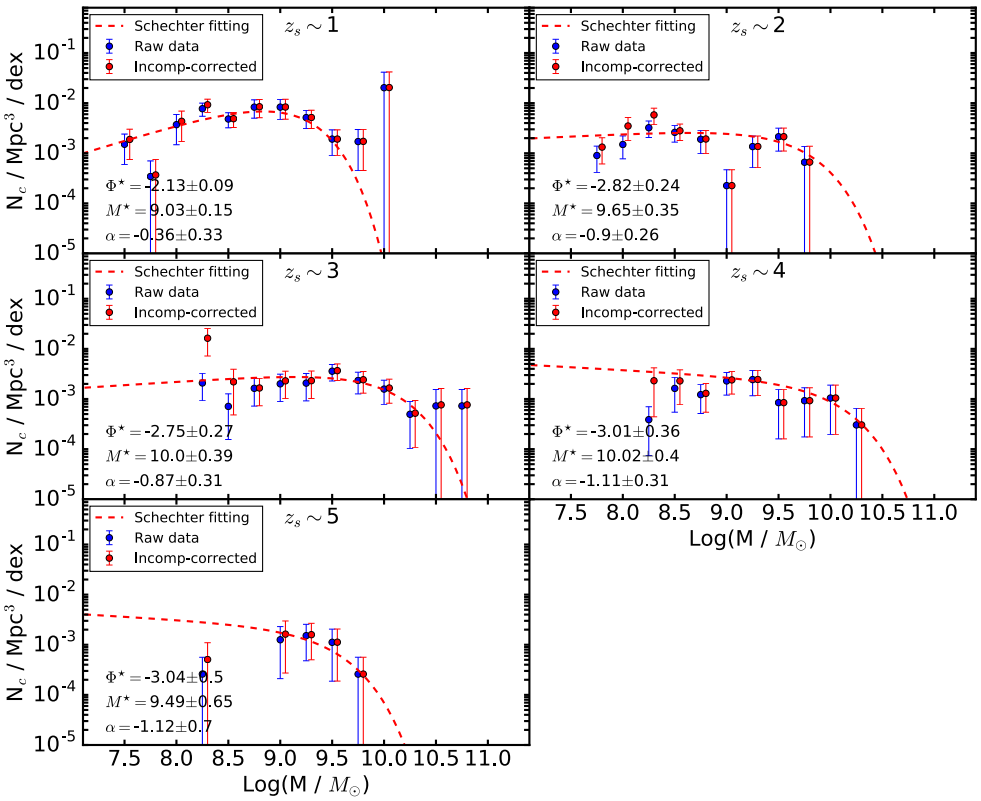


Figure 4.20: The incompleteness corrected stellar mass function of clumps at 5 redshift bins. The blue points are raw data and red points are the incompleteness corrected data. The red dashed lines are the best-fitted function by Schechter parameterization.

progenitors of the today's M31-mass galaxies (Papovich et al., 2015) shown by abundance matching. The stellar mass function of bulges show good agreement with that of clumps at high redshifts ($z_s \sim 3,4$) for M $> 10^{9.5}$ M$_\odot$ within uncertainties, but declines slightly more rapidly than that of clumps at low mass end (M $< 10^8$ M$_\odot$). The most massive clumps at $z_s \sim 3,4$ are also rest-frame UV brightest based on our study of luminosity function, which implies their intense star-forming activity. On the other hand, intriguingly, in our study of fundamental plane of clumps, we have shown that the stellar mass, size and compactness of the most massive clumps are also comparable with the bulge systems. There is evidence that at least the most massive clumps at high redshifts could be the progenitors of the local galactic bulges, which are undergoing intense star formation to assemble their mass. To assess whether these observed massive clumps at high redshift are the products of the coalesce of smaller clumps via migration requires detailed information of radial distribution of clumps at different redshifts, which we will discuss in the next chapter.

4.10 Global Properties of Clumps and Connection With the Host Galaxies

There are in general two proposed scenarios concerning the origins of clumps: the in-situ origin in which the clumps form inside the host gas disk via VDI

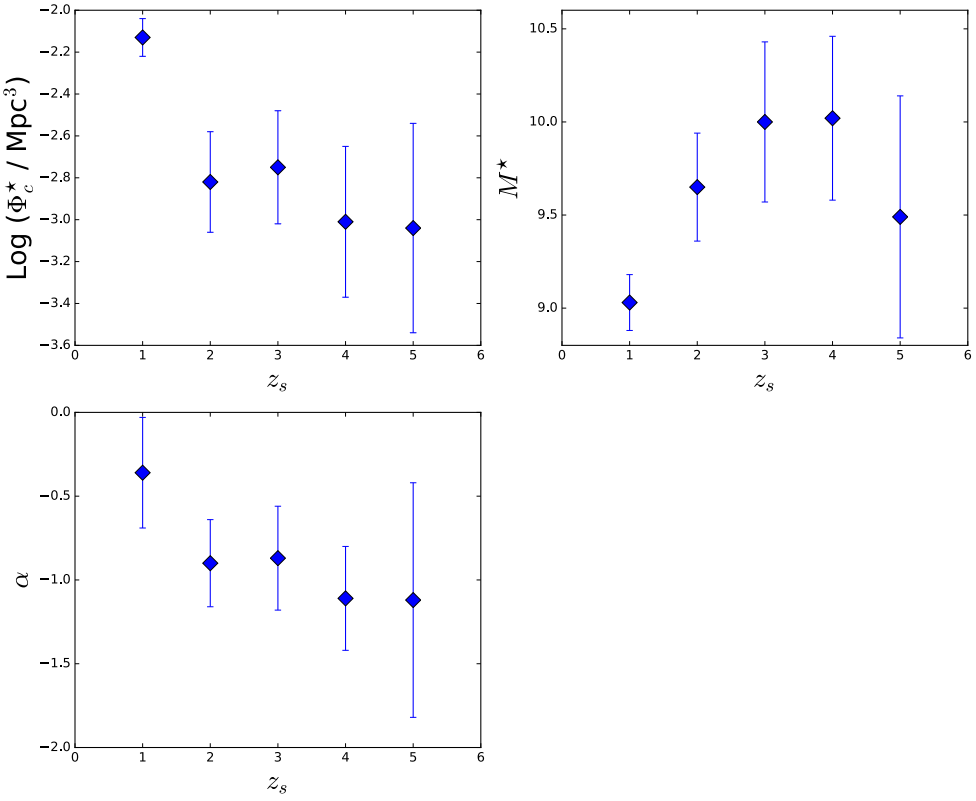


Figure 4.21: The redshift evolution of the best-fit Schechter parameter of stellar mass function. The error bar denotes 1σ rms error

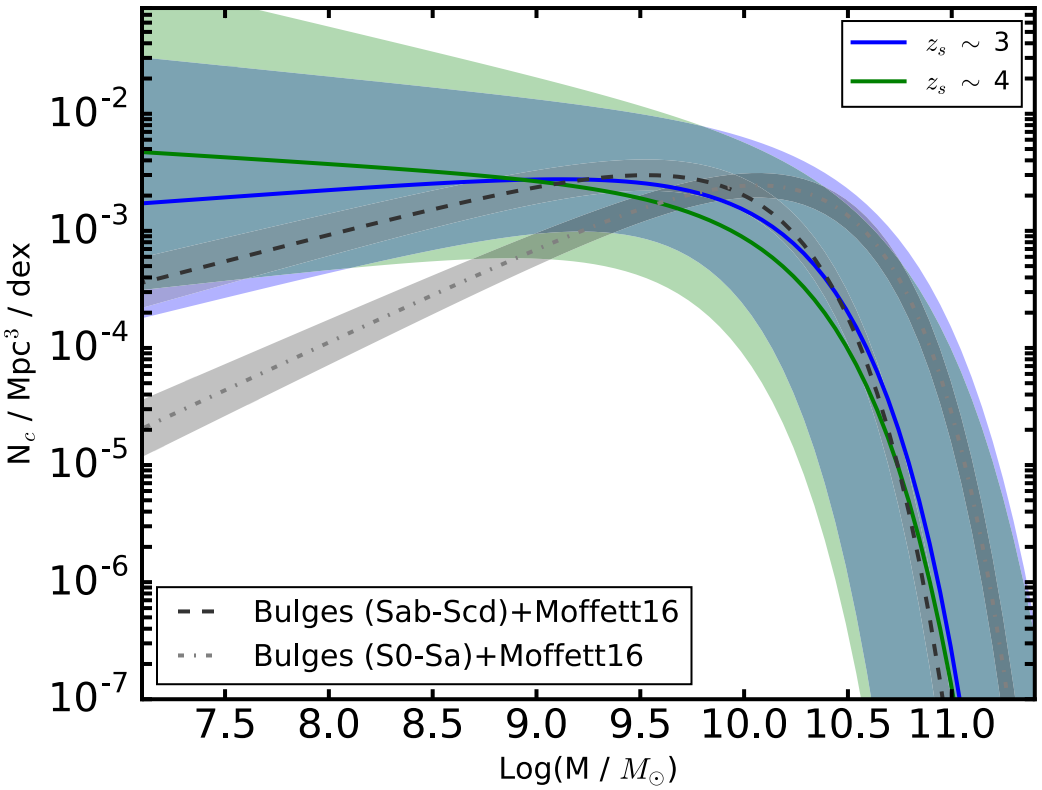


Figure 4.22: The stellar mass functions with best-fit Schechter parameters at $z_s \sim 3, 4$. The gray dashed and dot-dashed lines are the stellar mass function of local bulge in intermediate/late type and S0-Sa galaxies (Moffett et al., 2016).

(Violent Disk Instability); or ex-situ origin in which clumps are small dwarf galaxies that are captured by larger host galaxies. The relation between the clumps and host galaxies in two scenarios are different: the physical properties of ex-situ clumps is not expected to correlated with that of host galaxies since the capture of the infalling dwarf galaxies is random. Therefore, a detailed investigation of the connection between the clumps and host galaxies would help to clarify the formation mechanism of clumps. In the previous sections, we have shown evidence that the evolution of clumps is likely to correlate with that of the host galaxies. In this section, we mainly focus on two global properties: the multiplicity of clumps (number of clumps per galaxy) and the ratio of rest-frame UV luminosity and stellar mass of clumps to the host galaxies, to further explore the possible connections between the clumps and their host galaxies.

4.10.1 Multiplicity of Clumps

Figure 4.23 shows the distribution of the multiplicity of clumps (number of clumps per galaxy). We compare our multiplicity distribution with result from Elmegreen et al. (2013), in which 135 clumps were selected from rest-frame UV images of 30 clumpy UDF galaxies spanning the redshift range $1.1 < z_s < 3.6$. The AB limiting magnitude in UDF field for point source in all bands is $m_{AB} \sim 29$, about 1.5-2 mag deeper than that of CLASH survey.

But the actual limiting magnitude in survey area with high magnification is deeper than that in area with low magnification. The median magnification of CLASH lensed galaxy sample is ~ 5, which leads to a $\sim$ 2.5log10(5) = 1.7 mag deeper limiting magnitude, that is comparable with that in UDF field. In addition, their detection criteria of clumps appears to includes more fainter and less massive clumps with a median stellar mass of log($\mathrm{M}_\star/\mathrm{M}_\odot$) = 7.23, which is an order of magnitude smaller than the mass limit used in our CLASH study. However, the distribution of multiplicity appears to be insensitive to this large difference in stellar mass. A KS test on the distribution of multiplicity between their sample and our sample in same redshift range shows a p-value of 0.9998, and indicates that the distributions are drawn from the same parent distribution above a 5σ level.

We further test the dependence of multiplicity on the physical properties of host galaxies such as stellar mass, color and sSFR at different redshifts. In each test, we divide the full sample into two subsamples by the median value of the physical properties of the host galaxies, and compare the multiplicity distributions between the subsamples. Results are shown in Figure 4.24, 4.25 and 4.26, respectively. The KS tests show that we can not reject the null hypothesis that the subsamples divided by their host galaxy's properties are drawn from the same parent distribution in almost all the cases. The only exception occurs in subsamples divided by stellar mass of galaxies at $z_s \sim 5$, a

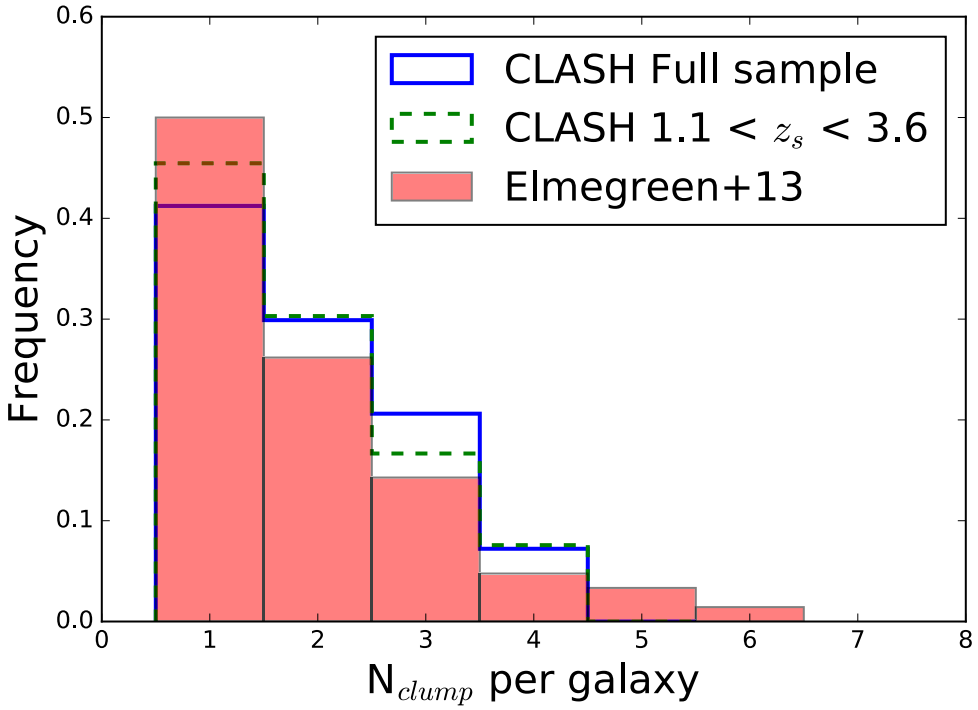


Figure 4.23: The distributions of multiplicity of clumps. The blue solid histogram is for the full CLASH sample; the green dashed histogram is for $1.1 < z_s < 3.6$ only; the red histogram is the distribution of sample from Elmegreen et al. (2013), which spans the redshift range $1.1 < z_s < 3.6$.

p value of 0.097 indicates that we can marginally reject the null hypothesis. We do not find any significant dependence of multiplicity on the physical properties of host galaxies. We conclude that the distribution of multiplicity is insensitive to the physical properties of the host galaxies.

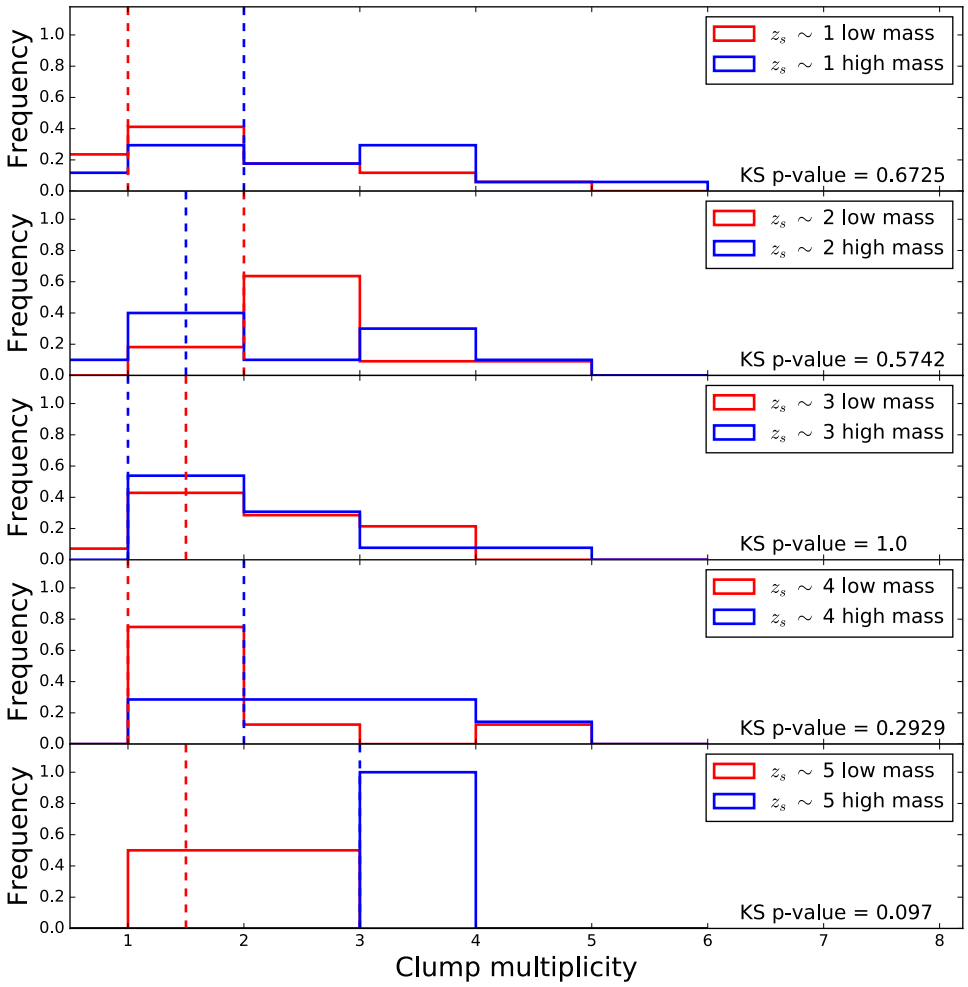


Figure 4.24: Comparison of the distributions between the subsamples of clumps divided by the median stellar mass of their host galaxies. The blue and red dashed lines are the median value of the distributions.

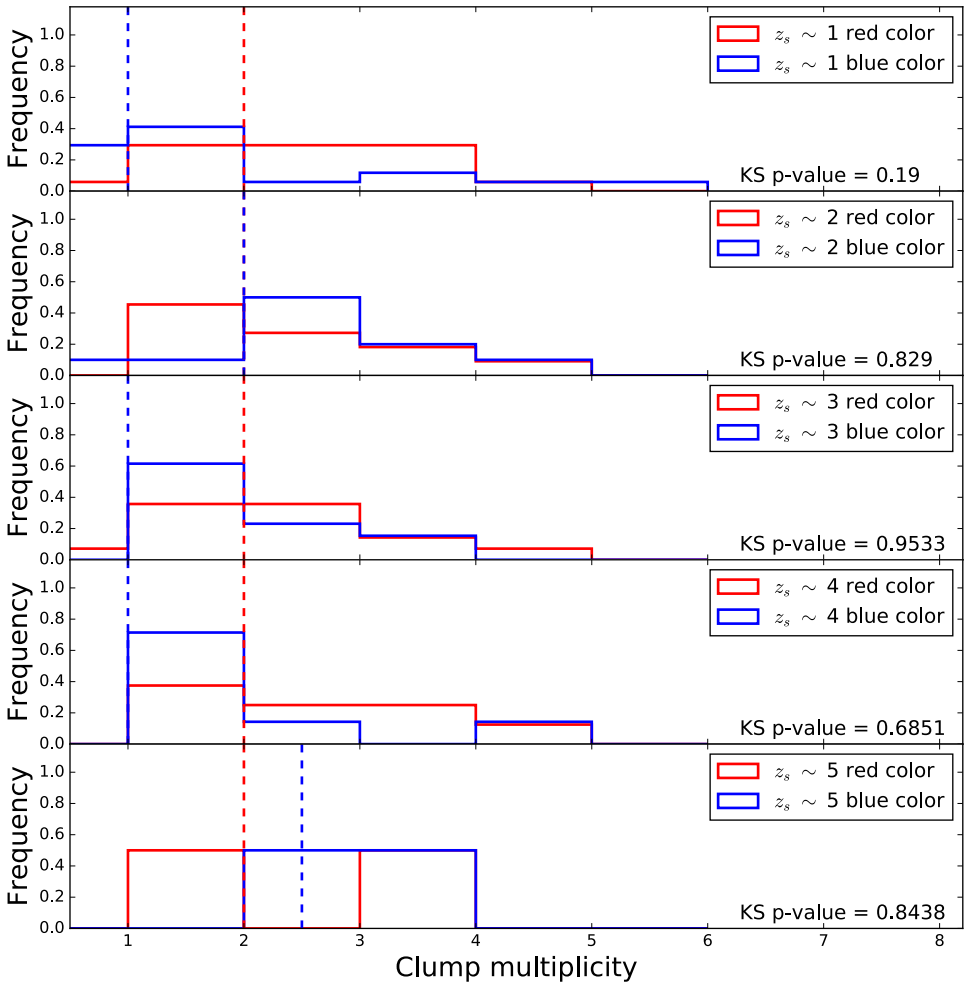


Figure 4.25: Comparison of the distributions between the subsamples of clumps divided by the median color of their host galaxies. The blue and red dashed lines are the median value of the distributions.

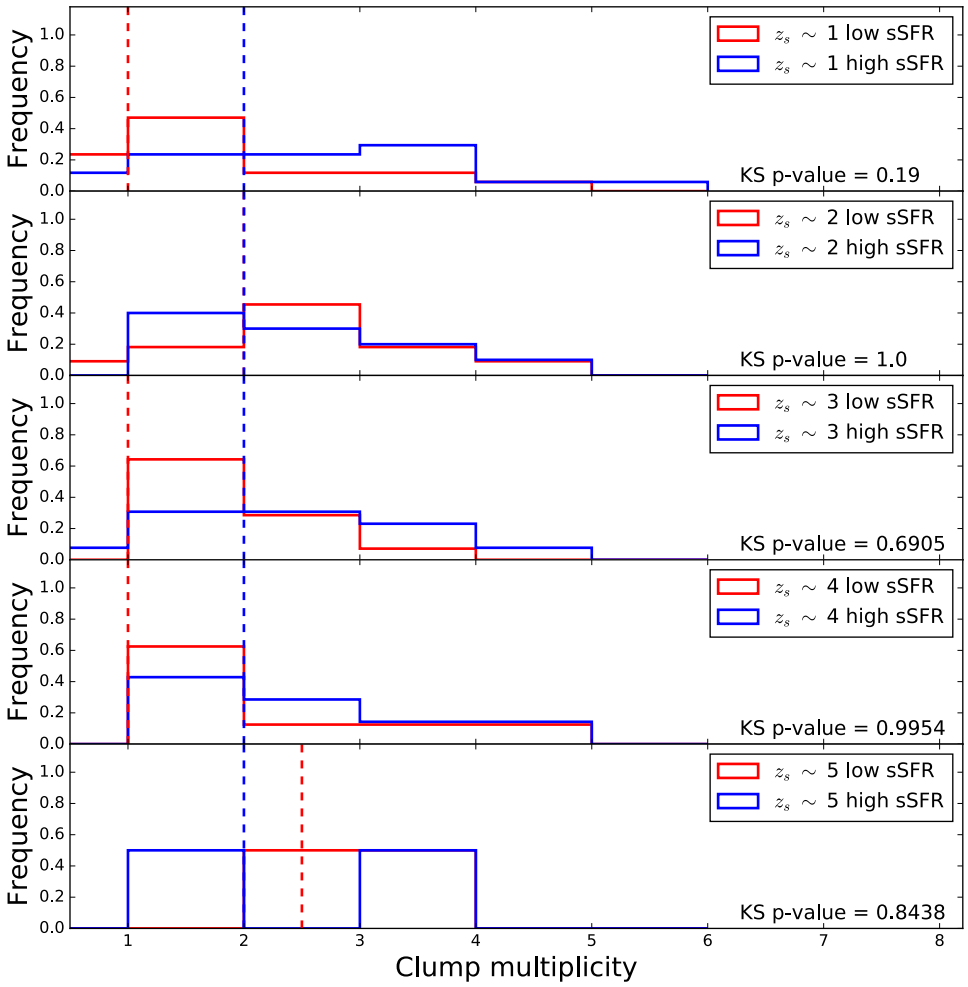
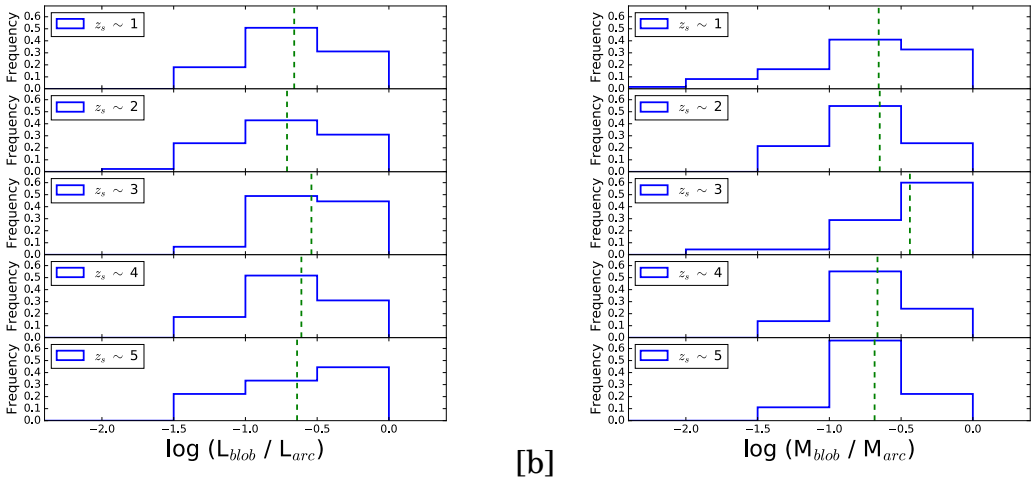


Figure 4.26: Comparison of the distributions between the subsamples of clumps divided by the median sSFR of their host galaxies. The blue and red dashed lines are the median value of the distributions.

4.10.2 Clump Contribution to the Light and Mass of Host Galaxies

We then study the contributions of the clumps to the total rest-frame UV and stellar mass of their host galaxies as a function of redshift (see Figure 4.27). The scatter and median value of distributions of UV contribution of clumps are similar at all redshifts. The scatter remains 1.5 dex and the median value is $\sim$ 10% - 20%, which is consistent with previous studies at $z_s < 3$ (Elmegreen et al., 2005, 2009; Forster Schreiber et al., 2011; Genzel et al., 2011; Guo et al., 2012; Wuyts et al., 2012). The median value of stellar mass ratio also remains constant over all the redshifts around 25% - 40%. The median value is however much higher than the typical value at lower redshift $\sim 5 - 10\%$ (Soto et al., 2017). This difference is likely due to a selection effect: our more stringent selection criteria preferentially selects brighter clumps. The median UV contribution also suggests that the UV-to-stellar mass ratio may be lower than the value at lower redshift, which is support by some observations (Grazian et al., 2015).

We show the evolution of the UV and stellar mass contributions in Figure 4.28. The average high median value of stellar mass contribution is consistent very well with that of identified "bulge" clumps in their nomenclature (central dense concentration) ~ -0.4 in numerical simulations (Mandelker et



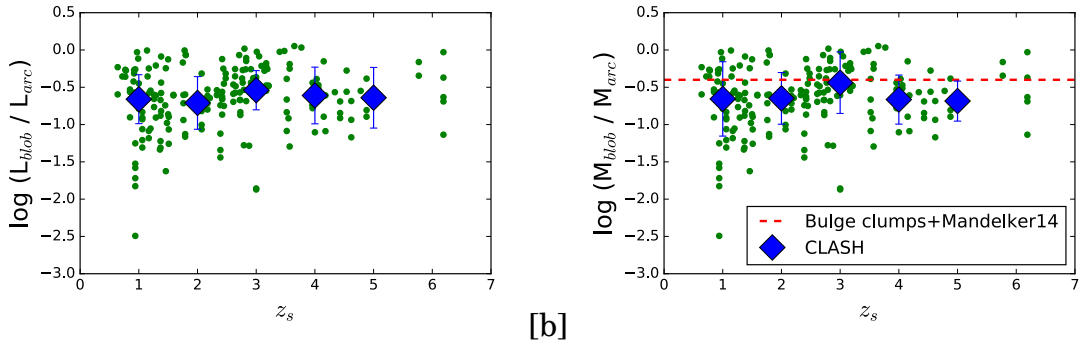
[a] [b]

Figure 4.27: The distributions of the rest-frame UV and stellar mass contribution of clumps to the host galaxies at 5 redshift bins. The green dashed lines are the median values of the distributions.

al., 2014, 2017), which further supports the hypothesis of a possible relation between the most massive clumps at high redshift to the local bulges. We also note that scatter in stellar mass contribution becomes larger at lower redshift, which implies that more smaller clumps with lower stellar mass contribution are identified at lower redshift than at higher redshift.

4.11 Summary and Conclusions

We have detcted 193 clumps from rest-frame 2200$\mathring{A}$ image of 106 CLASH lensed galaxies via devised clump-finder. We measure their physical properties such as stellar mass, SFR, sSFR, color and size. The main findings in this



[a] [b]

Figure 4.28: The redshift evolutions of rest-frame UV and stellar mass contribution of clumps to the host galaxies. The green dots are the CLASH data. The blue diamonds are the median value within the redshift bin and the error bar is the 1σ standard deviation. The red dashed line in (b) marks the typical stellar mass contribution value of bulge clumps that are identified from Mandelker et al. (2014)

chapter are summarized as follows:

1. The distribution of stellar mass of clumps correlates with that of the host galaxies. The dispersion in the distribution of various physical properties increases as redshift decreases, which implies the diversity in dynamical range and stellar population of clumps increase at lower redshifts.

2. The clumps at higher redshifts are more compact than those at lower redshifts. The evolution in compactness is insensitive to the selection of clumps at different scales.

3. The clumps fall onto the galaxies branch on the fundamental planes rather than branch of star cluster. The stellar mass, surface mass density and effective radius of clumps are very similar to those of dwarf spheroidals. The proposed origin of the dwarf spheroidals leads us to speculate that there could

also be evolutionary track linking the dwarf spheroidals and high-z clumpy galaxies.

4. The trend in characteristic magnitude and stellar mass of rest-frame UV luminosity function and stellar mass function of clumps correlates well with the trend in stellar mass of the host galaxies. The faint end slope of luminosity function shows weakly increasing trend as redshift decreases which appears to follow that of galaxies in high-z; the low mass end slope of stellar mass function remains flat except for $z_s \sim 1$, which implies that the clumps form in a scale independent way. The stellar mass function of clumps at high-z ($z_s \sim 3, 4$) is consistent with that of local late type galaxies at high mass end ($\mathrm{M} > 10^{9.5} M_{\odot}$), whereas the bulge mass function declines more rapidly than that of clumps at low mass end ($\mathrm{M} < 10^{8}\mathrm{M}_{\odot}$), which suggests the possible connection between the most massive high-z clumps and local bulge systems.

5. The multiplicity (number of clumps per galaxies) of clumps is insensitive to the physical properties of the host galaxies, and also not much affected by different selection in dynamical range.

6. The average rest-frame UV contribution of clumps to host galaxies is $\sim 10\% - 20\%$; The average stellar contribution of clumps to host galaxies is $\sim 25\% - 40\%$, which is much larger than the typical value at lower redshift. This value is consistent with that derived for bulge clumps (central dense concentration) in numerical simulations, which again suggests link between the

high-z clumps and local bulge systems.

Chapter 5

Spatial Distribution of Clumps

5.1 INTRODUCTION

The main epoch of the cosmic star formation occurs at redshift $z_s = 1 - 4$, when most of the stellar mass is assembled into galaxies (Madau et al., 1996; Hopkins & Beacom, 2006; Madau & Dickinson, 2014). Observations have shown that a fraction of at least $\sim 30\%$ galaxies at this epoch have rotating discs with high star formation rate (SFR) of order of 100 $\mathrm{M_\odot\ yr^{-1}}$ (Genzel et al., 2006; Forster Schreiber et al., 2009; Gnerucci et al., 2011). The CO measurement reveals that the gas fraction in high-z star-forming galaxies (SFGs) is $\sim 0.2 - 0.8$ (Daddi et al., 2010; Tacconi et al., 2013), which is much higher than the fraction in today's discs $\sim 0.05 - 0.1$ (Saintonge et al., 2011). The gas in the high-z SFGs is highly turbulent, with high velocity dispersion $\sim 20 - 80\ \mathrm{kms^{-1}}$

and low ratio of rotation to dispersion $V/\sigma \sim 2-7$, which is lower than that in today's spiral galaxies $\sim 10-20$ (Elmegreen et al., 2005; Genzel et al., 2006; Forster Schreiber et al., 2009). The high-z SFGs typically exhibit irregular and clumpy morphology in rest-frame UV and optical images (Elmegreen et al., 2005; Forster Schreiber et al., 2009), in the sense that $\sim 20-40\%$ of the UV light is concentrated in a few substructures – "clumps" with higher flux contrast (Guo et al., 2015) and typical size ~ 1 kpc or less (Elmegreen et al., 2005; Livermore et al., 2015). These clumps also contribute a few per cent of their disc mass and total SFR (Wuyts et al., 2012; Guo et al., 2015).

Given the clumpy morphology along with the underlying rotating discs and high gas fraction at high redshift, it has been long hypothesized that the origin of the high-z clumps are due to the violent disc instability (VDI; Dekel et al. (2009b)), which proposes that a Q parameter (see Section 4.1) that is lower than unity will make the gas disc unstable and fragment into large star-forming clumps, which has been widely shown in numerical simulations (Noguchi, 1999; Immeli et al., 2004a; Bournaud, Elmegreen & Elmegreen, 2007; Bournaud & Elmegreen, 2009; Hopkins et al., 2012). The losses of gas in forming stars will be replenished by the inflow of cold gas in narrow streams along the filaments of the cosmic web (Keres et al., 2005; Dekel & Birnboim, 2008; Dekel et al., 2009a), which maintains the high gas fraction and high surface density in the high-z SFGs.

Numerical simulations have shown that the gravitational torques can be natually induced by VDI to drive gas inflow towards the galactic center (Noguchi, 1999; Immeli et al., 2004a; Bournaud & Elmegreen, 2009; Zolotov et al., 2015). This dissipative "wet" inflow would cause central starburst that leads to "blue nuggets", which are compact, star-forming galaxies (Barro et al., 2013; Bruce et al., 2014). The intense stellar wind or AGN feedback will quickly quench the star formation in "blue nuggets", which then evolve into "red nuggets", which are compact, quiescent elliptical galaxies (Trujillo et al., 2007; Damjanov et al., 2009). Though "red nuggets" refers to massive galactic spheroidals with stellar mass $\sim 10^{11}\mathrm{M}_{\odot}$, it is likely that similar mechanism also applies to the formation of less massive spheroidals, say, as a dissipative growth mode of galactic bulges (Zolotov et al., 2015). Meanwhile, the gravitational torques as well as the dynamical friction could also drive the massive clumps migrate towards the galactic center, which contribute to the dry growth mode of the galactic bulges. In parallel with the major-merger driven bulge formation scenario in the classic models of galaxy formation (Milos & Hernquist, 1996; Hopkins et al., 2006), the VDI induced gas inflow and clumps migration offer an alternative scenario of the built-up of the galactic bulges, which does not require an frequent merger environment.

If the clumps are long-lived, they are supposed to migrate towards the galactic center in a few dynamical timescales at the disc edge (Dekel et al., 2009b;

Mandelker et al., 2014). Hence, a key issue that clump migration scenario needs to confront is: can clumps survive the stellar feedback in the process of migration? The estimated lifetime of clumps by numerical simulations roughly fall into two categories: simulations that only consider the feedback from supernova conclude that all the clumps can be long-lived, with a lifetime $\sim$ 250 - 500 Myr (Dekel et al., 2009b; Bournaud et al., 2014; Mandelker et al., 2014); whereas in simulations that consider more efficient radiation trapping and momentum driven feedback, all the clumps are quickly disruptd by feedback with a much shorter lifetime $\sim$ 25 - 50 Myr (Genel et al., 2012; Hopkins et al., 2012; Oklopcic et al., 2017). Mandelker et al. (2017) implements moderate radiative feedback along with the feedback from supernova in simulations of clumps and find that the number of long-lived clumps are significantly reduced due to the inclusion of the radiative feedback, and only more massive and compact clumps can survive the radiative feedback as expected. To get rid of the ambiguity on the model-dependent lifetime of clumps, observational constraints are urgently needed to not only clarify the longevity of the clumps, but also to infer the dominant feedback mechanism for clumps, as well as series of other implications based on VDI scenario.

There have been several observational measurements on stellar age of the clumps (Elmegreen et al., 2005, 2009; Forster Schreiber et al., 2011; Genzel et al., 2011; Wuyts et al., 2012; Guo et al., 2012), which range from 10 Myr to

a few Gyr. The large age spread as well as the large uncertainties limits our ability to draw direct conclusions on the lifetime of clumps. Moreover, as the clumps migrate inwards, they will continue to accrete fresh gas to form new stars. These newly-born young stellar populations would likely dominate the UV flux budget therefore potentially bias the age measurement. On the other hand, some of these studies have attempted to infer the migration of clumps based on the observed radial gradients of clump properties (Forster Schreiber et al., 2011; Guo et al., 2012). e.g. Guo et al. (2012) find that clumps that are closer to the galactic center tend to have redder color, older age, lower specific SFR (sSFR). However, the strength to distinguish different model predictions based on solely the radial gradients within narrow redshift range appears to be limited. e.g. an inside-out growth model could also reproduce similar color gradient of clumps (Wuyts et al., 2012); Oklopcic et al. (2017) find that even short-lived clumps that barely migrate can reproduce various similar radial trends as Guo et al. (2012) identifies. Therefore, new observables that can both constrain the lifetime of the clumps and tie to the unique features of migration scenario are not currently available.

We study the radial distribution of clumps as well as the radial variation of clump properties such as stellar mass, surface mass density, sSFR and color to test the longevity of clumps and clump migration scenario, based on the sample of 193 clumps that are described in Chapter 4. The sample spans the

redshift range $1 < z_s < 6$, which for the first time, allows us to investigate the evolution of the radial distribution of clumps. We suggest that the evolution of the radial distribution is an unique tool to disentangle the model prediction of clump migration from predictions of other models, thus is crucial to access the validity of the VDI scenario.

5.2 Measuring the Radial Distribution of Clumps

5.2.1 The Centroids of the Lensed Galaxies and Clumps

To study the projected distance of clumps, we first need to determine where the galactic center is. Unlike most of today's galaxies which have definite concentration of mass or flux in the innermost region, the star-forming galaxies at $z_s > 1$ typically have irregular morphology and do not have well defined center of mass or flux. We make use of the spatially-resolved stellar mass map to compute the mass-weighted center $\vec{R_c}$:

$$\vec{R_c} = \frac{\sum_i m_i \vec{r_i}}{\sum_i m_i}, \tag{5.1}$$

where i is the index of pixels that belong to the reconstructed image in the source plane; m_i and $\vec{r_i}$ are the mass and position of the pixel i, respectively. We define this center of mass $\vec{R_c}$ as the centroid of each lensed galaxy.

For each detected clump, we locate the pixel with maximum rest-frame UV intensity within the aperture of clump as the centroid of clump $\vec{r_c}$. The absolute projected pixel distance as well as the angular and physical distance are then computed as the distance between the centroid of lensed galaxy and clump

$$d_{proj} = ||\vec{R_c} - \vec{r_c}||. \tag{5.2}$$

5.2.2 Normalized Projected Distance of the Clumps

To compare the projected distance of clumps in different galaxies, we normalize the projected distance by the semi-major axis of the host galactic disc. To estimate the length of semi-major axis of the galactic disk, we first compute the entire stellar mass by summing all the pixels enclosed by the segmentation map of galaxies in the spatially-resolved stellar mass maps. We then construct elliptical apertures centered on the mass-weighted center in the spatially-resolved stellar mass maps. We vary the semi-major axis of the ellipse until the aperture encloses the entire mass budget, and the final semi-major axis is our estimated galactic radius R_d. Using the normalized projected distance has the advantage of being less affected by the bias or inaccuracy from

the lensing mass models: both the absolute distance and galactic radius are affected by the inaccuracy of the lensing mass models to the same extent, hence the ratio is less biased than the absolute value.

The measured normalized projected distances of clumps d_{proj}/R_d are listed in Table 4.1.

5.3 Quantifying the Radial Dependence of Clumps

5.3.1 Parameterization of the Radial Dependence of Clumps

We now describe the methodology to quantify the radial variation of the physical properties of clumps. The measurements of physical properties of clumps such as stellar mass, color, SFR and sSFR are provided in Chapter 4. To study the radial evolution of clumps, we first divided the clump sample into 5 redshift bins. In each redshift bin, we assign the clumps into 5 radial bins by their normalized projected distance. The radial bin size is determined by the maximum and minimum projected distance in each redshift bin. The median value and 1σ standard deviation are calculated in each radial bin. We then fit

the radial dependence of clumps based on the following parametric form:

$$\mathrm{PARAM} \propto (\mathrm{d_{proj}}/\mathrm{R_d})^{\alpha}, \tag{5.3}$$

$$\alpha = \frac{\mathrm{dlog(PARAM)}}{\mathrm{dlog(d_{proj}/R_d)}}, \tag{5.4}$$

where PARAM is the value of physical parameters, α is logarithmic slope of the radial dependence. We also compute the Spearman rank correlation coefficient and its p-value to test the significance of the correlation of data in each redshift bin.

5.3.2 Uncertainties

To take the uncertainties from estimation of photometric redshift and SED fitting of clumps into account, we conduct 200 Monte Carlo simulations on the catalog by perturbing the value of photometric redshift and physical parameters within 1σ errors that are estimated from BPZ and iSEDfit. In each realization, we reassign the clumps into 5 redshift bins by the perturbed redshift and re-calculate the median value and 1σ standard deviation of the physical parameters in each radial bin. In performing the radial dependence fitting, we run bootstrapping for 200 times to randomize the radial distance data. The 1σ scatter in the resulting radial dependence is then our total uncertainty, which

is:

$$\sigma_{\rm tot} = \sqrt{\sigma^2_{\rm photoz} + \sigma^2_{\rm SED} + \sigma^2_{\rm bootstrapping}}. \quad (5.5)$$

In addition, we calculate the Spearman rank correlation coefficient and the related p-value in each redshift bin to test the statistical significance of the radial gradients.

5.4 RESULTS

5.4.1 Radial Distribution of Clumps

We first compare the radial distribution of clumps at different redshifts. Since the migration of clumps is believed to be mainly driven by the gravitational torques of the host galaxies, it is natural to investigate the migration of clumps from less and more massive galaxies separately, because the stellar mass is the most important dynamical property of the host galaxy. We divide the full clump sample into two subsamples by the median stellar mass of the host galaxy $\mathrm{M}_\star^{\rm med} = 10^{9.5}\ \mathrm{M}_\odot$ (hereafter, we name them high mass and low mass samples). Figure 5.1 shows the radial distributions of the whole sample and two subsamples as a function of redshift. Due to a lack of less massive clumps in $z_s \sim 5$ bin, we only plot the distribution of full and high mass sample at $z_s \sim 5$. The radial distribution of the full sample does not change at differ-

ent redshifts. The median projected distance of the full sample stays around $d_{proj} \sim 0.3R_d$ over all the redshifts. The median radial distance in low mass sample show weakly deceasing trend as redshift deceases. The median radial distance in high mass sample decreases slightly until $z_s \sim 3$ and then increases towards the lower redshift. We then compute the median radial distance and 1σ standard deviation in each redshift bin and fit the slopes $\alpha = d(d_{proj}/R_d)/dt$ for high and low mass samples, as Figure 5.3 shown. Comparing with the predicted slope $\alpha \sim 2$, 1 and 0.5, our slopes for both samples ($\alpha_h = 0.003 \pm 0.058$ and $\alpha_l = 0.04 \pm 0.063$) are consistent with no radial evolution or very weak evolution, which is potentially in conflict with the expectations from the clump migration scenario.

As discussed in introduction, the clumps are expected in simulations to form in the outer region of the massive disc and migrate towards the galactocenter driven by the gravitational torque and dynamical friction. Therefore, instead of indirectly inferring the clump migration from the radial gradient of physical properties, we argue that the most straightforward and indisputable observable to test clump migration is *whether the radial distribution of clumps is moving closer to the galactocenter as the redshift decreases*. The inside-out growth model would predict an evolution trend in which clumps will generally be moving further away from the galactocenter as redshift decreases if the clumps can form in the growing disc. If the clumps do not migrate at all or mi-

grate very slowly, the radial distribution should remain unchanged as redshift varies.

However, the above finding does not allow us to rule out the clump migration model. First of all, all the prediction for clump migration are based on very massive simulated disk galaxies which typically have stellar mass $> 10^{10}\mathrm{M}_{\odot}$; whereas our lensing selected sample mainly includes smaller galaxies (e.g. we only have 3 galaxies which have stellar mass greater than $3 \times 10^{10}\mathrm{M}_{\odot}$). So the only conclusion we can draw here is that our data is more consistent with that the clump does not migration or migrate very slowly in smaller galaxies. Second, star-forming galaxies at $z_s > 1$ with irregular morphology typically do not have well defined galactocenter, and the galactocenter defined in our study is the center of the stellar mass. If the clumps mass dominates the total mass of the host galaxies, this definition would make the more massive clumps tend to lie closer to the galactocenter, thus even an observed significant radial trend does not necessarily correspond to the migration of clumps. Therefore, firm confirmation or falsification of clump migration scenario requires a robust comparison between observation and simulation that share the same definition of radial distance and host galaxies with similar mass range. Further larger clump samples are needed to revisit the scenario of clump migration in massive galaxies.

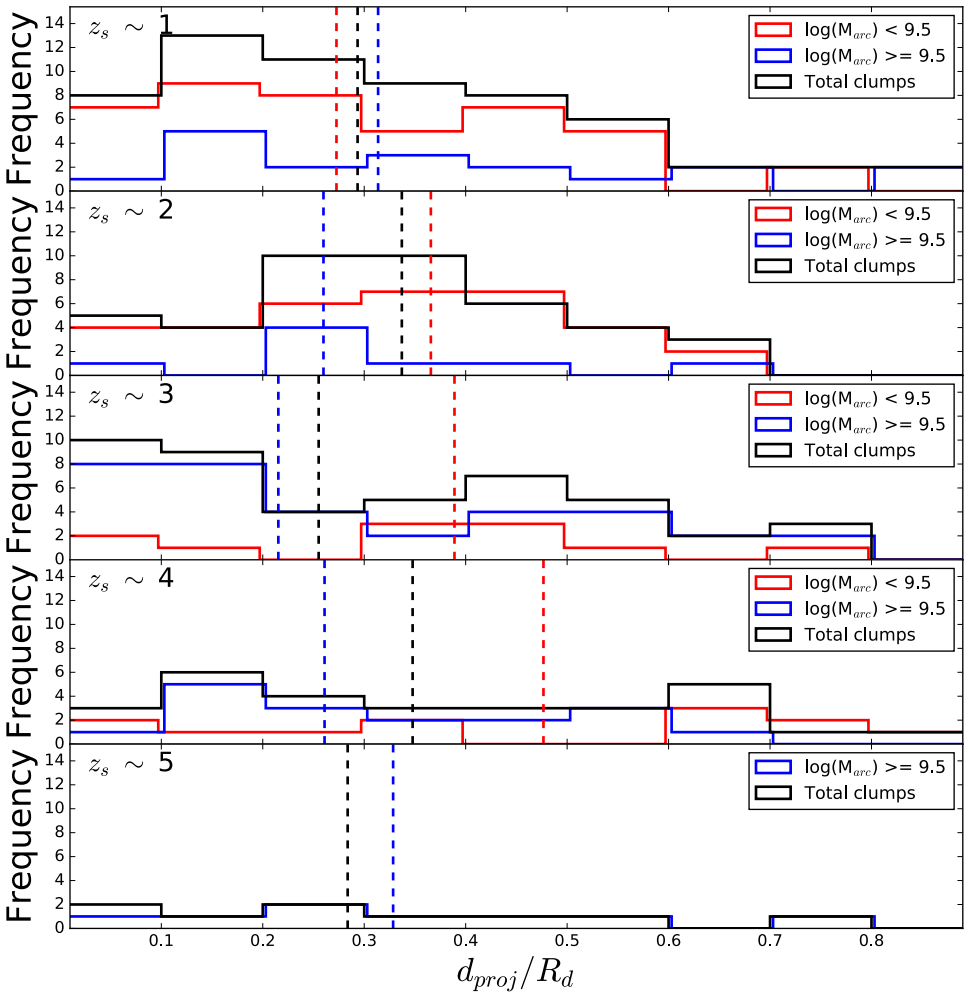


Figure 5.1: The radial distributions of full sample of clumps and two subsamples divided by the median stellar mass of the host galaxy at 5 redshifts. The colored dashed lines denotes the median value of different distributions.

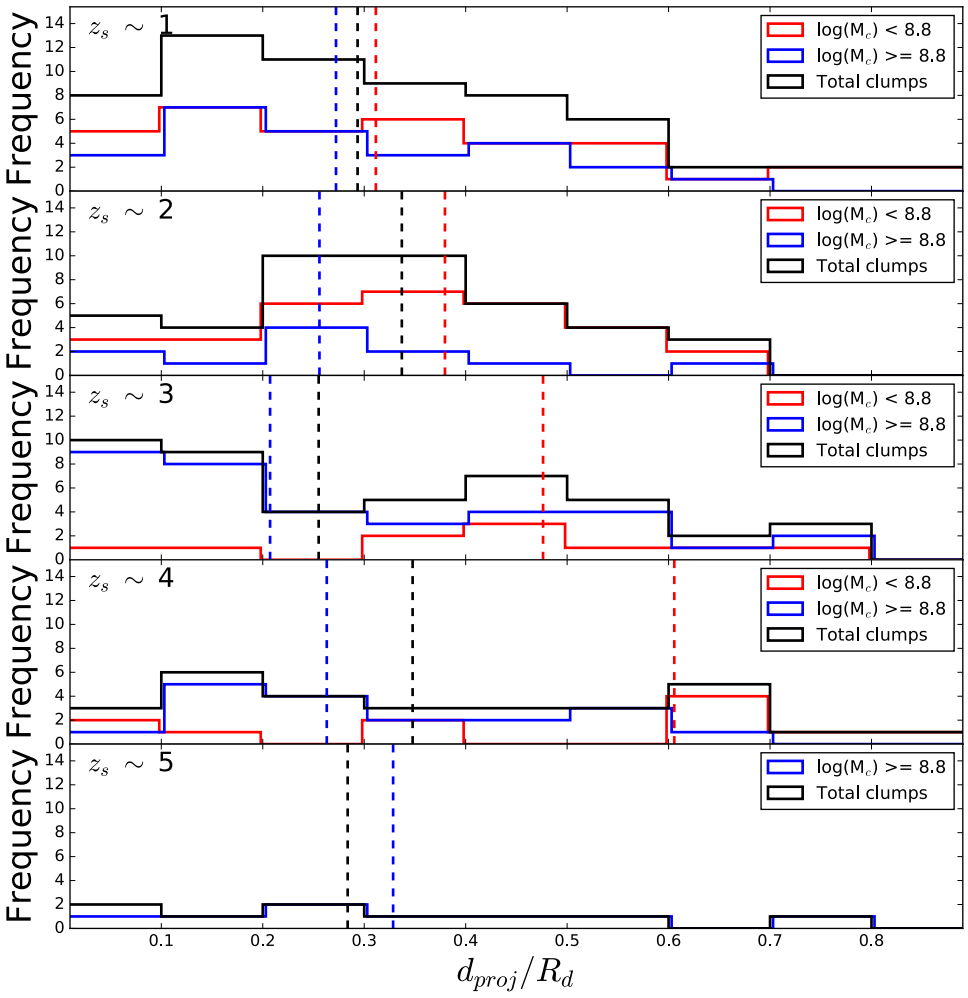


Figure 5.2: The radial distributions of full sample of clumps and two subsamples divided by the median stellar mass of clumps at 5 redshifts. The colored dashed lines denotes the median value of different distributions.

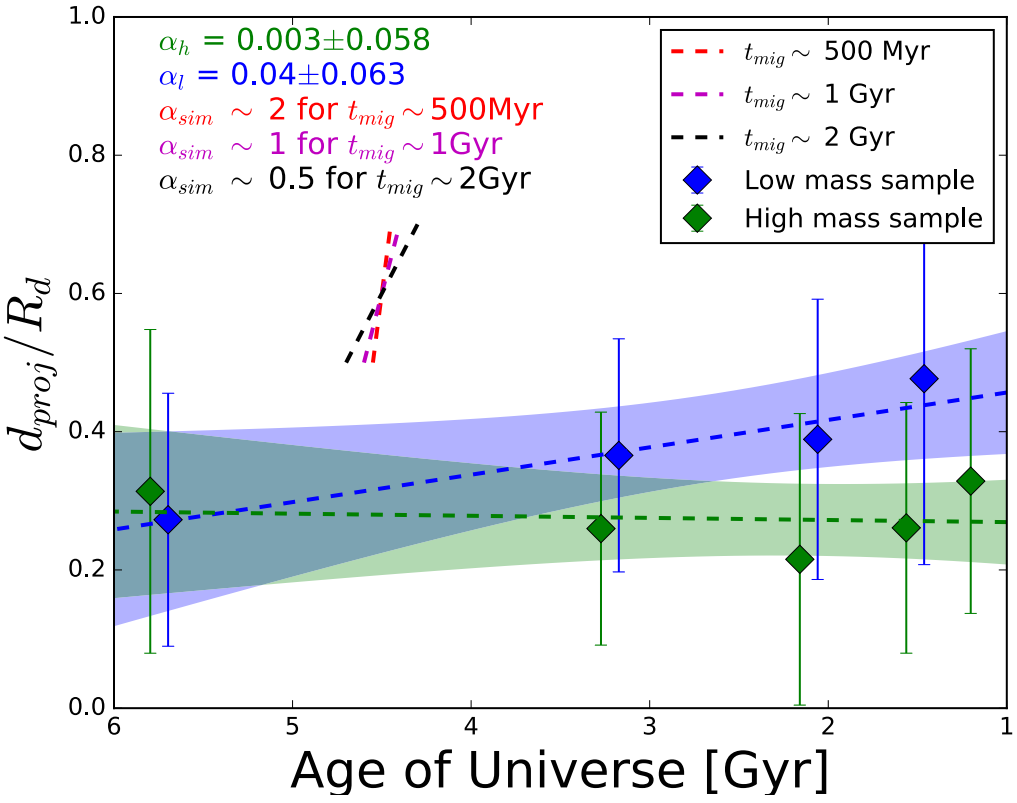


Figure 5.3: The radial evolution of clumps in high and low mass subsamples at 5 redshifts. The green (blue) diamond and error bars denote the median value and 1σ standard deviation in each redshift bin for high (low) sample. α_h and α_l are the best fitted slopes of radial evolution for high and low mass subsamples and the color shaded regions are 1σ scatter of the best fitted lines. The red, cyan and black dashed lines denote the predicted slopes of radial evolution for timescale of migration $t_{mig} \sim$ 500 Myr, 1 Gyr and 2 Gyr.

5.4.2 Radial Variation of Physical Properties of Clumps: Full Sample

In this section, we investigate the radial variation of the physical properties of clumps for the full sample. Four properties of clumps, which are crucial to assess the longevity of clumps, have been explored: stellar mass, surface mass density, color and sSFR. Figure 5.4, 5.5, 5.6, 5.7 show the radial distributions of these properties for the full sample. We observe significant steep radial gradient $\alpha = -1.195 \pm 0.258$ in stellar mass at $z_s \sim 3$, and -1.107 ± 0.333 at $z_s \sim 4$. The p-values = 0.029 and 0.023 indicate that the correlation at $z_s \sim 3, 4$ are statistically significant. We observe steep slope 0.983 ± 0.482 at 2σ level in color and -0.622 ± 0.22 at 3σ level in surface mass density with marginal significant correlation at $z_s \sim 3$. Other than those, the radial distribution of clumps at all redshifts is consistent with flat. At $2.5 < z_s < 5.5$, the dispersion of all the properties increases as the redshift decreases. The dispersion of stellar mass, surface mass density and color reach $\sim$ 2 dex at $z_s \sim 3$ while the dispersion in sSFR reaches $\sim$ 1 dex. The scatters become slightly larger towards the lower redshifts. In addition, the median color for high mass sample increases $\sim$ 1 dex from $z_s \sim 3$ to $z_s \sim 1$ though the uncertainty is large. The increasing scatters in physical properties indicate that the diversity of clumps in stellar mass and stellar population properties increases,which is evidence that supports the

long-lived clumps because the increasing dispersion in color and sSFR are not expected in short-lived scenario. However, the small scatter at very high redshift could also be due to the the incompleteness. A larger sample of clumps at very high redshift would be useful to draw more solid conclusions on the lifetime of clumps.

5.4.3 Radial Variation Dependence on Multiplicity of Clumps

In Section 5.4.1, we emphasize that a galactocenter defined as center of stellar mass likely bias the projected distance of massive clump short for high-z irregular galaxies. In this section, we will assess whether this geometric effect is important in our radial measurement, and we will also discuss how this effect relates the multiplicity of clumps to their radial distribution for high-z irregular galaxies.

In massive galaxies which have well defined central concentration, the radial distribution of clumps should be irrelevant to the multiplicity of clumps. However, it might not be the case for high-z irregular galaxies without settled central concentration, in which the fractional mass of clumps is higher than their local counterparts. e.g. if a high-z irregular galaxy hosts one massive clump, we would expect that *in galaxies with similar stellar mass, the defined*

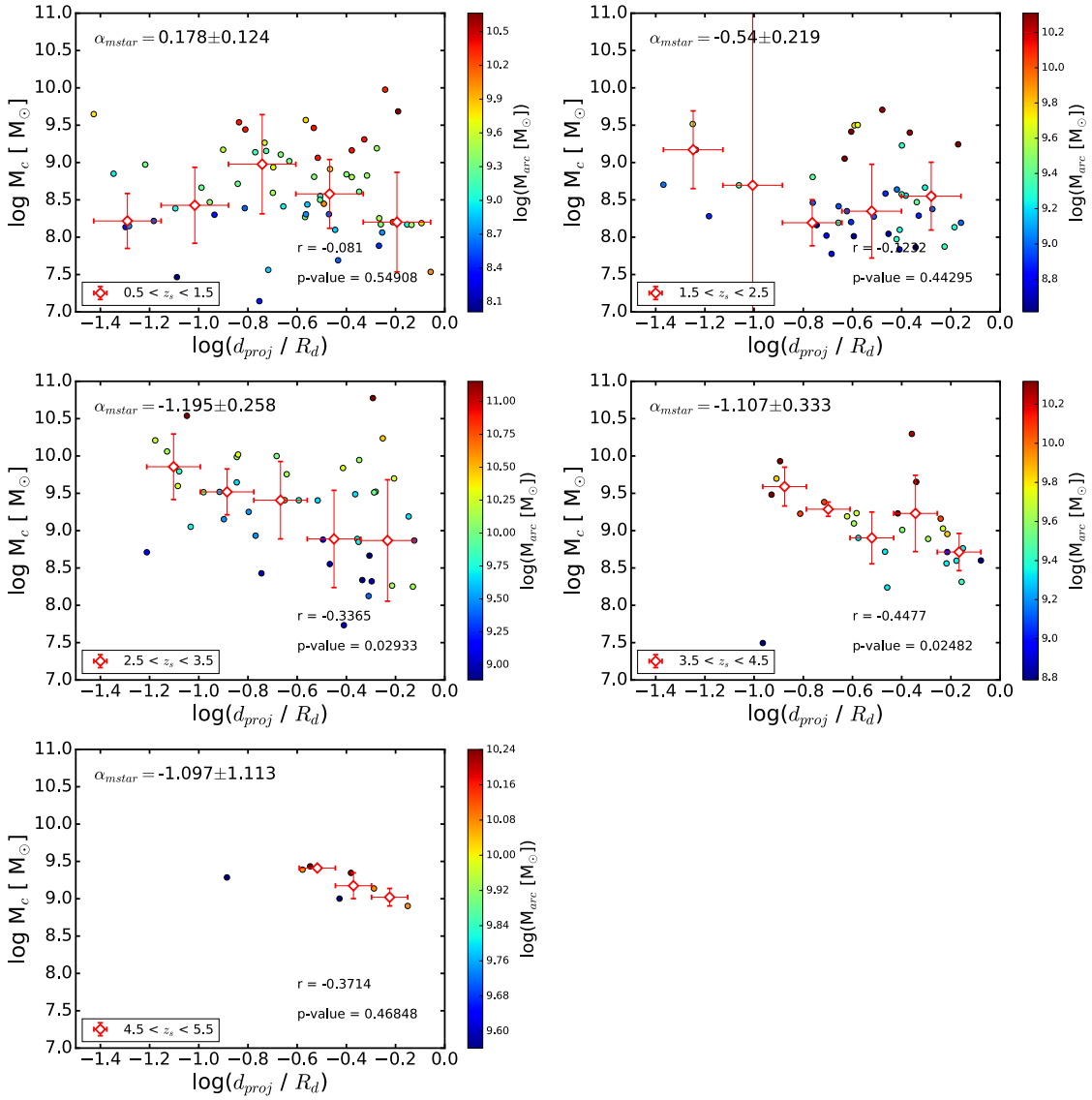


Figure 5.4: The radial variation of stellar mass of clumps at 5 redshifts, color-coded by their host galaxy mass. The diamond points and error bar denote the median value and 1σ standard deviation in each radial bin. The left-top is the logarithmic slope of the gradient; the right-bottom shows the Spearman rank correlation coefficient and the p-value.

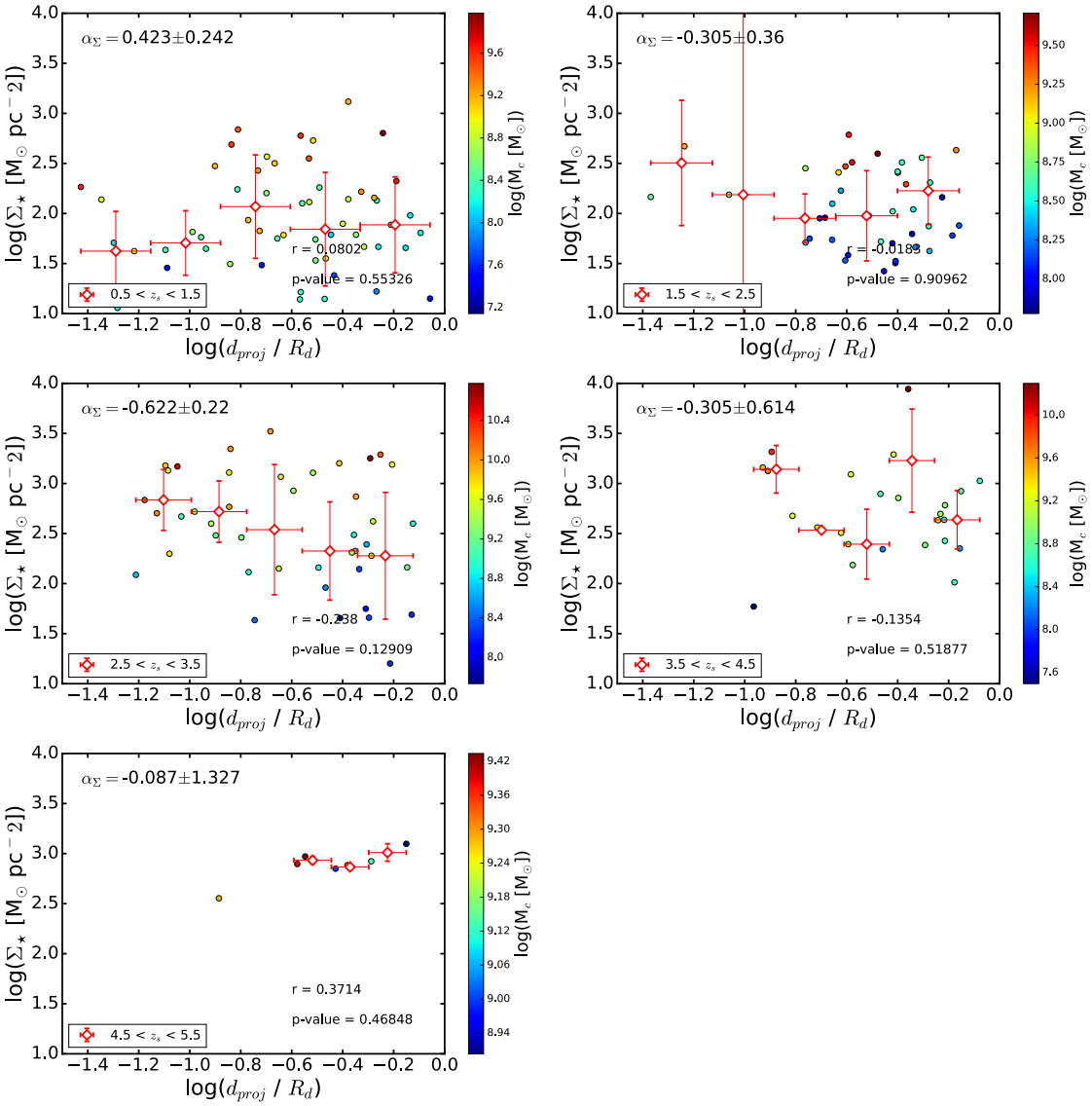


Figure 5.5: The radial variation of surface mass density of clumps at 5 redshifts, color-coded by the clump mass. The diamond points and error bar denote the median value and 1σ standard deviation in each radial bin. The left-top is the logarithmic slope of the gradient; the right-bottom shows the Spearman rank correlation coefficient and the p-value.

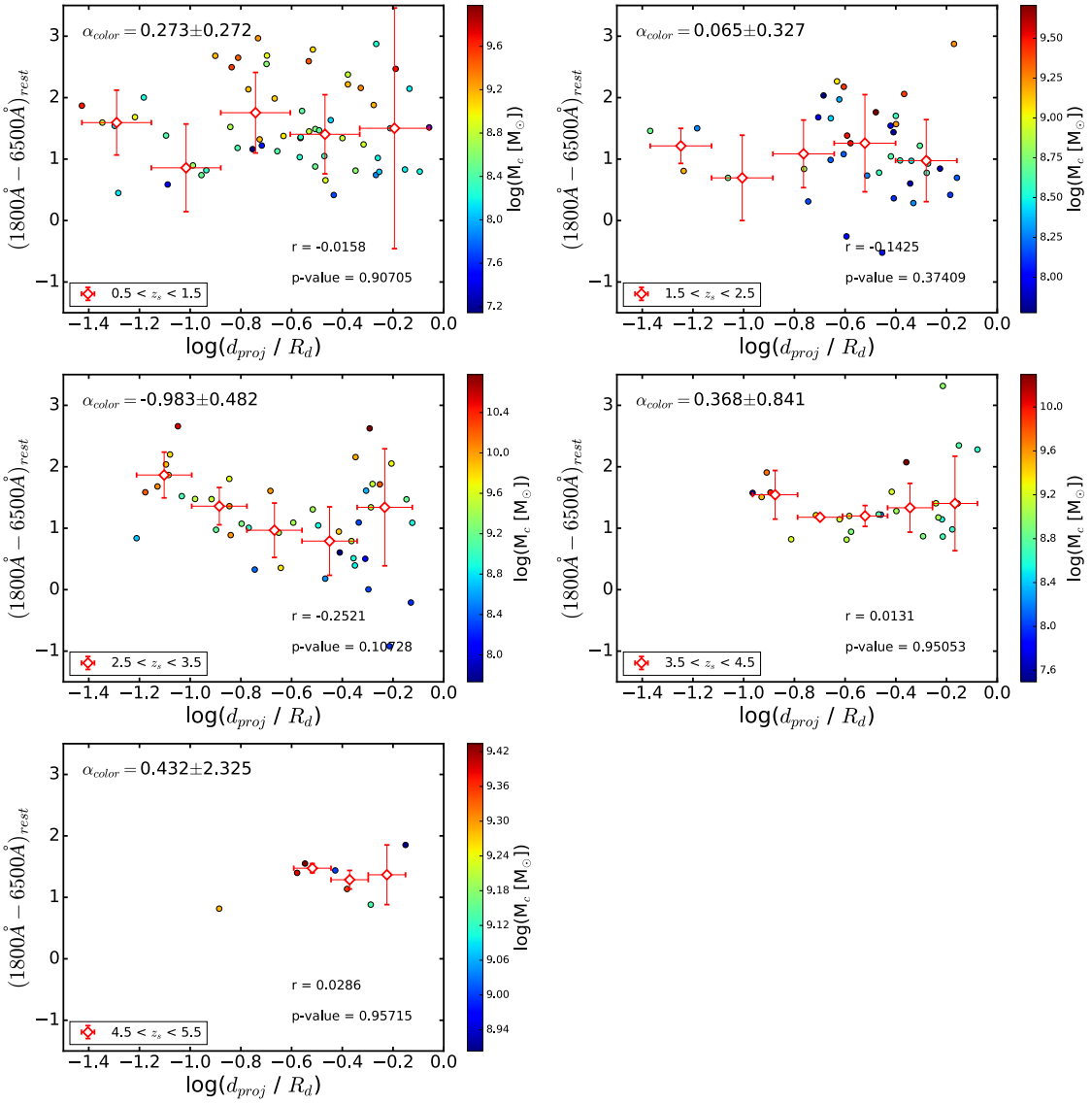


Figure 5.6: The radial variation of color of clumps at 5 redshifts, color-coded by the clump mass. The diamond points and error bar denote the median value and 1σ standard deviation in each radial bin. The left-top is the logarithmic slope of the gradient; the right-bottom shows the Spearman rank correlation coefficient and the p-value.

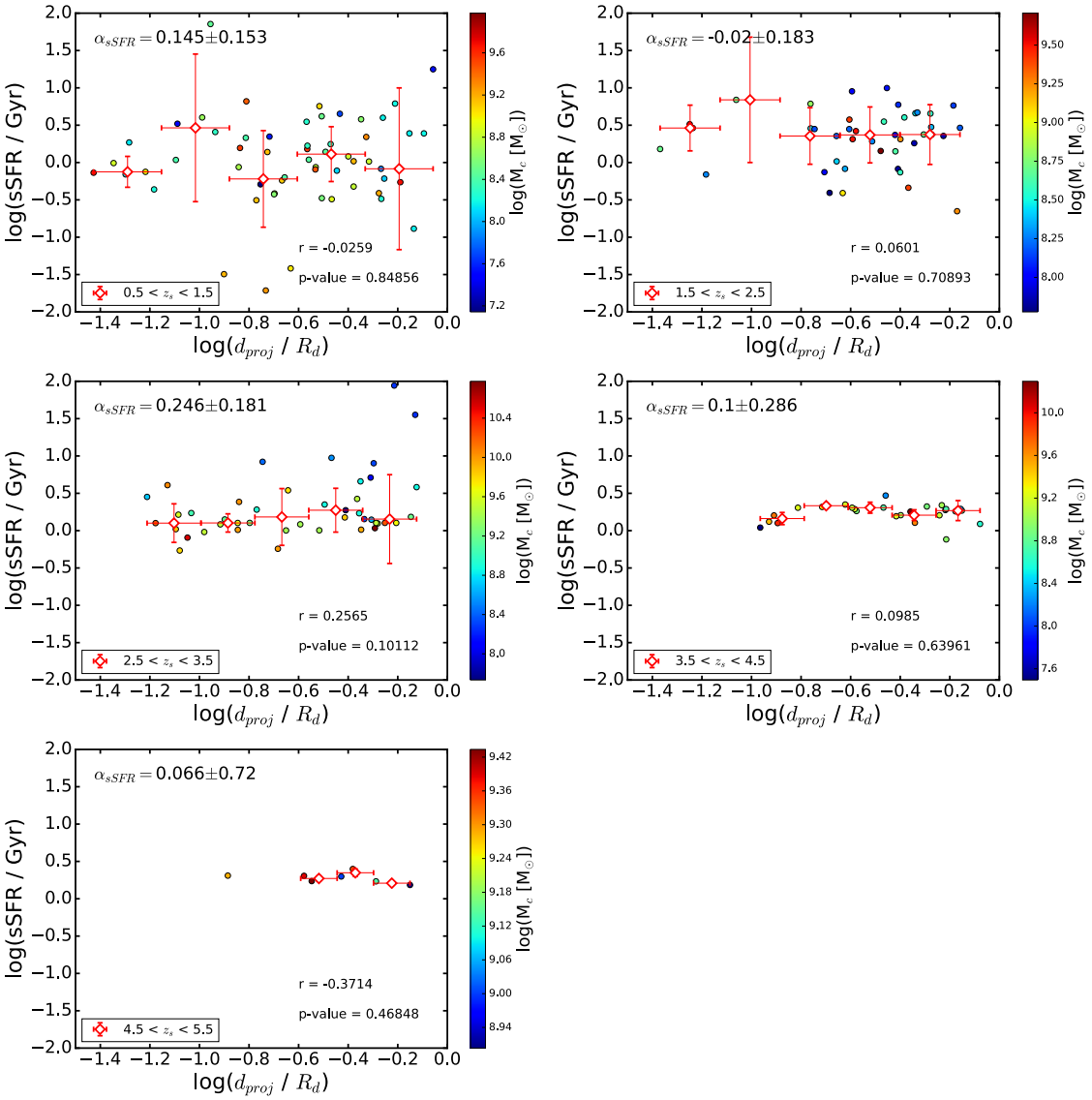


Figure 5.7: The radial variation of sSFR of clumps at 5 redshifts, color-coded by the clump mass. The diamond points and error bar denote the median value and 1σ standard deviation in each radial bin. The left-top is the logarithmic slope of the gradient; the right-bottom shows the Spearman rank correlation coefficient and the p-value.

barycenters of mass should be statistically closer those massive clump than the less massive ones. If the geometric effect is important, we would expect to see that the median projected distance of clumps in irregular galaxies that host single massive clump should be statistically larger than that in galaxies that host multiple clumps. To test this, we further construct four subsamples by dividing the high and low mass samples by the multiplicity of clumps: galaxies with only one clump and galaxies with multiple clumps in which the ratio of the most massive clump to the second most massive clump is smaller than 3:1. We plot the stellar mass of the most massive clump in each galaxy as a function of the projected distance of the most massive clump at 4 different redshifts (Figure 5.8). At all redshifts, almost all the clumps from galaxies with multiple clumps in both high and low mass samples stay at outer region of the disc with $\mathrm{d_{proj}} \geq 0.5\mathrm{R_d}$. We have the most data of single and multiple clumps in low mass sample at $z_s \sim 1$ and in high mass sample at $z_s \sim 3$. In low mass sample at $z_s \sim 1$, the median projected distance of single clumps of -0.77 ± 0.32 is larger than that of multiple clumps -0.33 ± 0.13 at 1.3σ level; while in high mass sample at $z_s \sim 3$, the median projected distance of single clumps of -0.84 ± 0.32 is larger than that of multiple clumps -0.29 ± 0.18 at 1.5σ level. Since only marginal difference has been detected, our conclusion that the clumps does not migrate or migrate very slowly in smaller galaxies appears not significantly affected by this geometric effect.

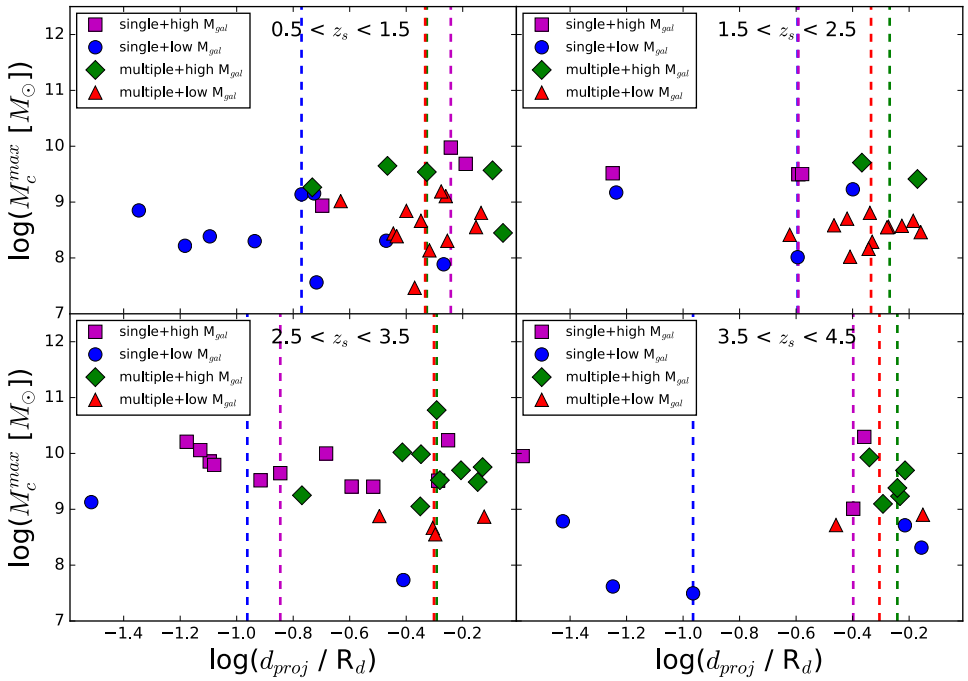


Figure 5.8: The stellar mass of the most massive clump in each galaxy as a function of its projected distance. Clumps from four subsamples are listed for comparison. "single+high" denotes the galaxy that has mass $>= 10^{9.5}\mathrm{M}_{\odot}$ and hosts single massive clump; "single+low" denotes the galaxy that has mass $< 10^{9.5}\mathrm{M}_{\odot}$ and hosts single massive clump; "multi+ high" denotes the galaxy that has mass $>= 10^{9.5}\mathrm{M}_{\odot}$ and hosts multiple massive clump; "multi+low" denotes the galaxy that has mass $< 10^{9.5}\mathrm{M}_{\odot}$ and hosts multiple massive clump. For galaxy with multiple clumps, only those with the mass ratio of the most massive clump to the second most massive clump smaller than 3 are selected. The colored dashed lines are the median value of the different samples.

However, the bias introduced by the geometric effect turns still a caveat which hampers us to draw robust conclusions on comparing the observed and simulated radial distributions, especially at high redshift. In simulation, the galactocenter is often pre-defined as the mass-weighted center of gas disc before the clump fragmentation. Thus the different way of defining the galactocenter before and after the clump fragmentation would bring in bias in the measurement of projected distance of clumps. Either adopting a definition of galactocenter in simulations based on the instantaneous mass distribution at each simulated snapshot, or using a independent definition of galactocenter in observations such as center of dynamical mass would help to reduce the bias in radial measurement.

5.5 Comparison with Simulations

As summarized in the introduction, the lifetime of clumps predicted by the numerical simulations is highly model dependent. Theoretic models based on different stellar feedback recipe would give different lifetime of clumps, ranging from 50 Myr - 500 Myr. In this section, we compare the observed results with results from two simulations and attempt to provide constraints on the clump formation model. One simulation for comparison has investigated the long-lived clumps (Mandelker et al., 2014, 2017), in which most clumps are

detected from host galaxies with median stellar mass $\sim 10^{10}\mathrm{M}_{\odot}$ at $1 < z_s < 2$, and their clump mass completeness limit is $\sim 10^{7}\mathrm{M}_{\odot}$. The other simulation has explored the short-lived clumps (Oklopcic et al., 2017), in which all the clumps are identified from a massive, gas-rich disky galaxy ($\mathrm{M}_{\star} \sim 10^{10.8}\mathrm{M}_{\odot}$ at $z_s \sim 1$) with a mass completeness limit $\sim 10^{7.5}\mathrm{M}_{\odot}$. They also explore the evolution of properties of clumps from $z_s \sim 2$ to $z_s \sim 1$. At similar redshift range, our clumps are generally detected from smaller host galaxies. The median stellar mass of host galaxy is $\sim 10^{9.3}\mathrm{M}_{\odot}$ for the full sample and $\sim 10^{10}\mathrm{M}_{\odot}$ for the high mass sample. Our clump mass completeness limit at $1 < z_s < 2$ is $\sim 10^{8}\mathrm{M}_{\odot}$, which is 0.5-1 dex higher than those of simulations.

5.5.1 Comparison of Stellar Mass Distribution of Clumps

The mass range of host galaxy for our high mass sample is close to that of Mandelker et al. (2017), so we first compute the incompleteness corrected stellar mass function of clumps at $1 < z_s < 2$ from high mass sample and compare with two simulated mass distributions of clumps with different stellar feedback implemented: with and without radiative feedback. The comparison is shown in Figure 5.9. Within the uncertainty, our low mass end slope of stellar mass function $\alpha = -0.78 \pm 0.36$ is consistent with the model which imple-

ments the radiative feedback that efficiently disrupt the less massive clumps ($\alpha \sim -1$). The slope would be steeper ($\alpha \sim -1.5$) if solely the feedback from supernova is considered. We also compute the incompleteness corrected fractional mass function at three redshift bins as Figure 5.10 shown. The trend is remarkably consistent with that in Figure 9 in Mandelker et al. (2017): the less massive clumps become more dominant as redshift decreases, and the peak fraction value slightly decreases as redshift decreases, which is due to that the disc masses grow monotonically whereas the mass completeness limit remains fixed. Note that our fractional mass is systematically larger than their value by ~ 2 dex. This is due to the combination of effects that we include the most massive in our sample that identified as "bulge" clumps in Mandelker et al. (2017), which is usually an order of magnitude more massive than the normal off-center clumps; and our completeness limit is 1 dex higher than theirs thus including less clumps with lower fractional mass that pushes the peak towards the right.

On the other hand, strong momentum-driven feedback is implemented in Oklopcic et al. (2017), which introduces a different style of death for clumps: the more massive clumps appear to be more efficiently disrupted. This can be clearly seen from their Figure 4., in which they compare the mass distribution of clumps at three redshifts $1.0 < z_s < 1.4$, $1.4 < z_s < 1.8$ and $1.8 < z_s < 2.2$. As redshift decreases to $1.0 < z_s < 1.4$, almost all the clumps with $\mathrm{M}_\star > 10^8 \mathrm{M}_\odot$

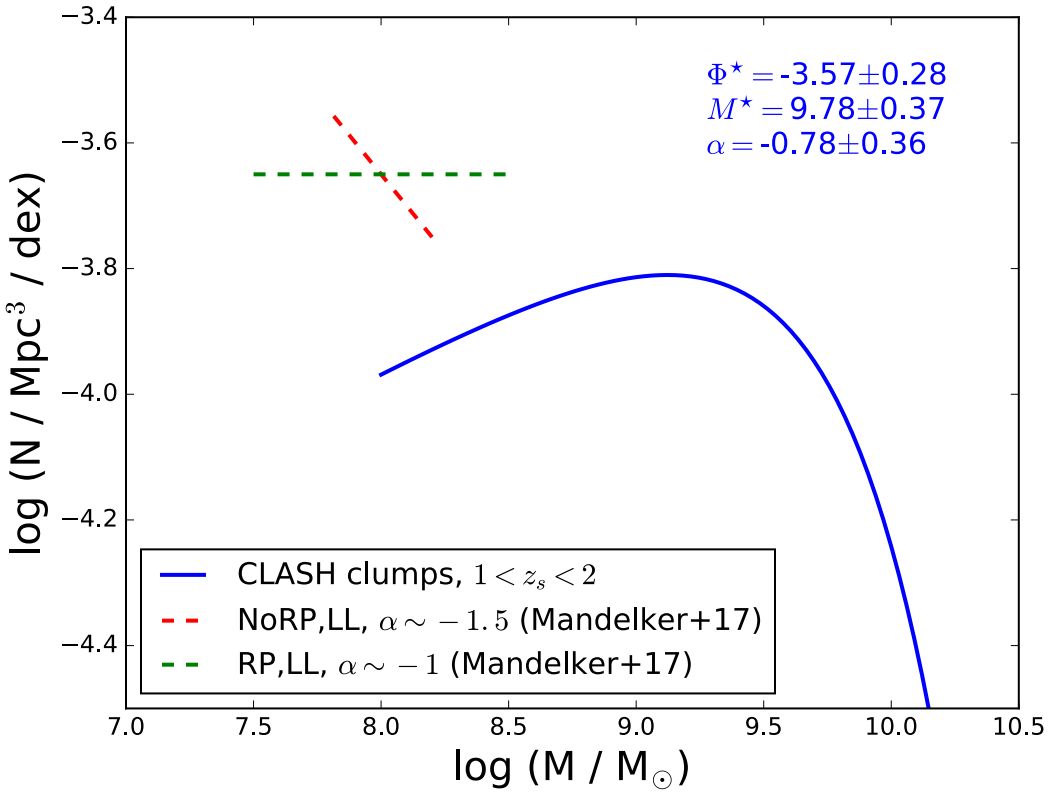


Figure 5.9: The comparison between the incompleteness corrected mass distribution of CLASH clumps at $1 < z_s < 2$ and the simulated results. The red dashed line denotes the low mass end slope of the simulated mass distribution without radiative feedback implemented; the green dashed line denotes the simulated low mass end slope with radiative feedback implemented. The simulated results are from Mandelker et al. (2017)

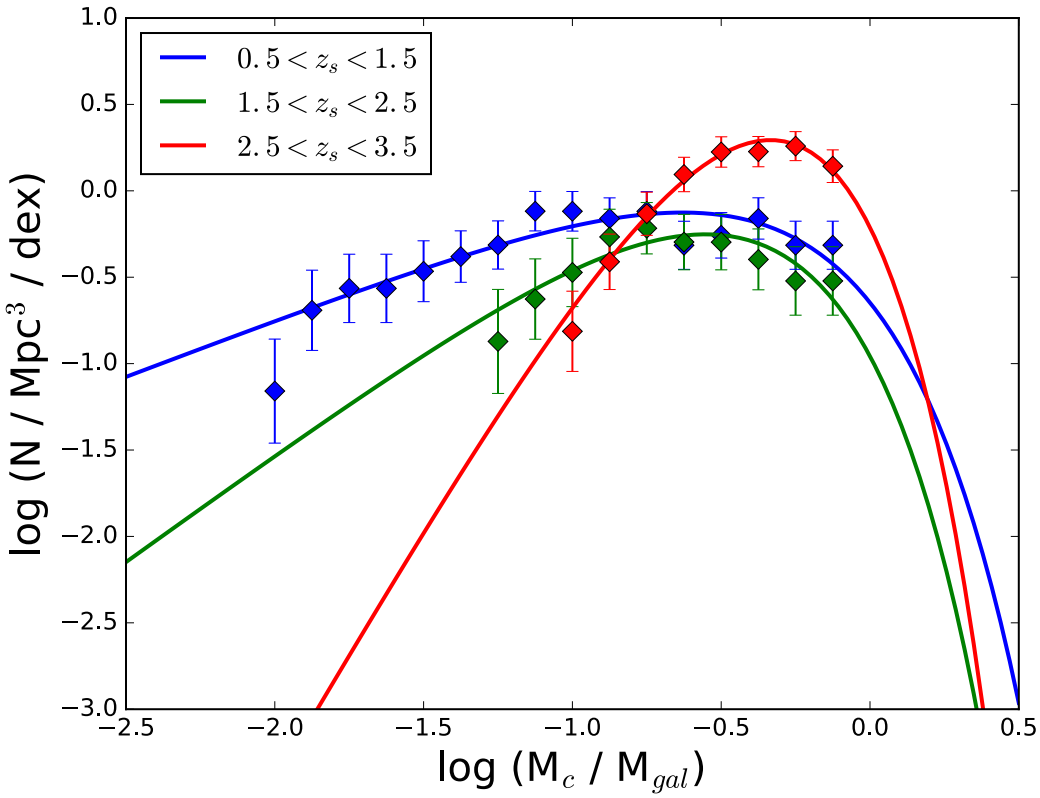


Figure 5.10: The incompleteness corrected fractional mass function of CLASH clumps at three redshifts. The colored diamonds are the incompleteness corrected CLASH data.

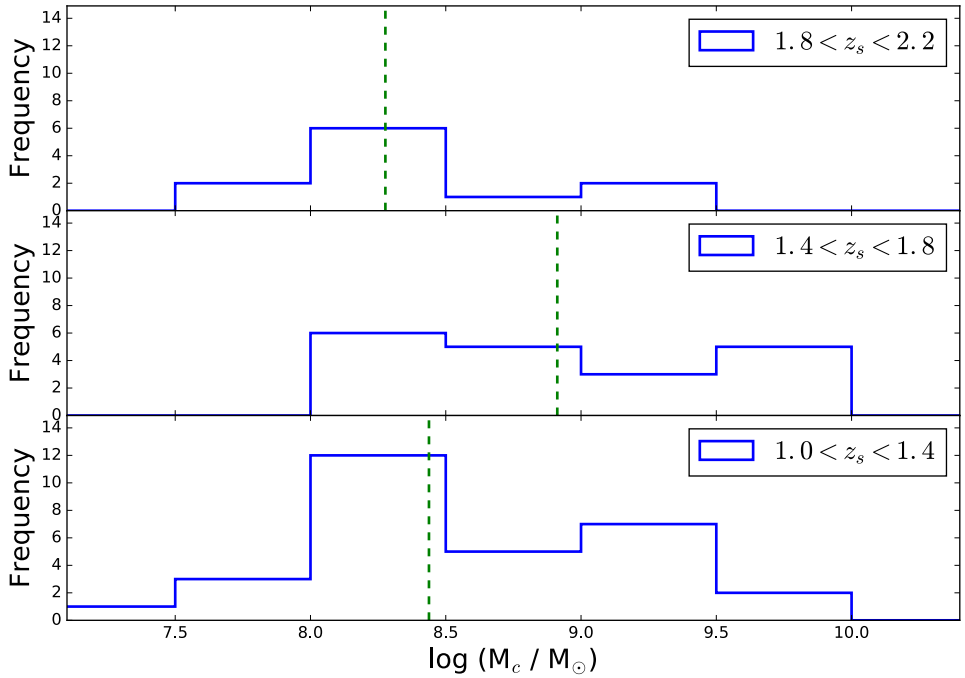


Figure 5.11: The stellar mass distribution of CLASH clumps at three redshift ranges. The redshift ranges are chosen as same as those from Oklopcic et al. (2017) for comparison.

have been disrupted. We plot our stellar mass distribution of clumps at same three redshifts in Figure 5.11. We do not detect clear deficit of clumps with $M_\star > 10^8 M_\odot$ as redshift decreases. This null detection combined with the consistency with Mandelker et al. (2017) demonstrates that the stellar feedback tends to disrupt the less massive clumps more efficiently, and the momentum-driven feedback appears not strong enough to quickly disrupt the more massive ones.

5.5.2 Comparison in Fundamental Plane of Clumps

Whether the clumps can survive the stellar feedback is closely related their stellar mass and compactness. Different predictions on the lifetime of clumps have implies that the simulated clumps in different simulations might be dynamically distinct. Therefore, before drawing conclusions on the lifetime of clumps from the comparison between the simulations and observations, we first need to answer an important question: *Do simulated clumps indeed have similar properties as the observed CLASH clumps?* As discussed in Section 3.7, the position in fundamental plane is a very sensitive diagnostic of the dynamical properties of stellar systems. Hence we plot the simulated long-lived and short-lived clumps in the fundamental plane of $\Sigma_\star - M_\star$ and compare with the observed clumps in Figure 5.12. It appears that both the long-lived and short-lived clumps are systematically more compact than the those detected in this CLASH study. The long-lived clumps cover similar mass range with the observed ones while the short-lived clumps overlap the low-mass end of the CLASH clump distribution. From their positions, the simulated long-lived clumps are dynamically more consistent with the observed clumps though $\sim$ 0.5 dex offset in $\Sigma_\star$ still exists to the main branch of real clumps; whereas the short-lived clumps appear to lie between the observed clumps and globular clusters. We suggest that future studies should attempt to confirm the dynamical similarity of the clumps with that of the observed clumps, prior to

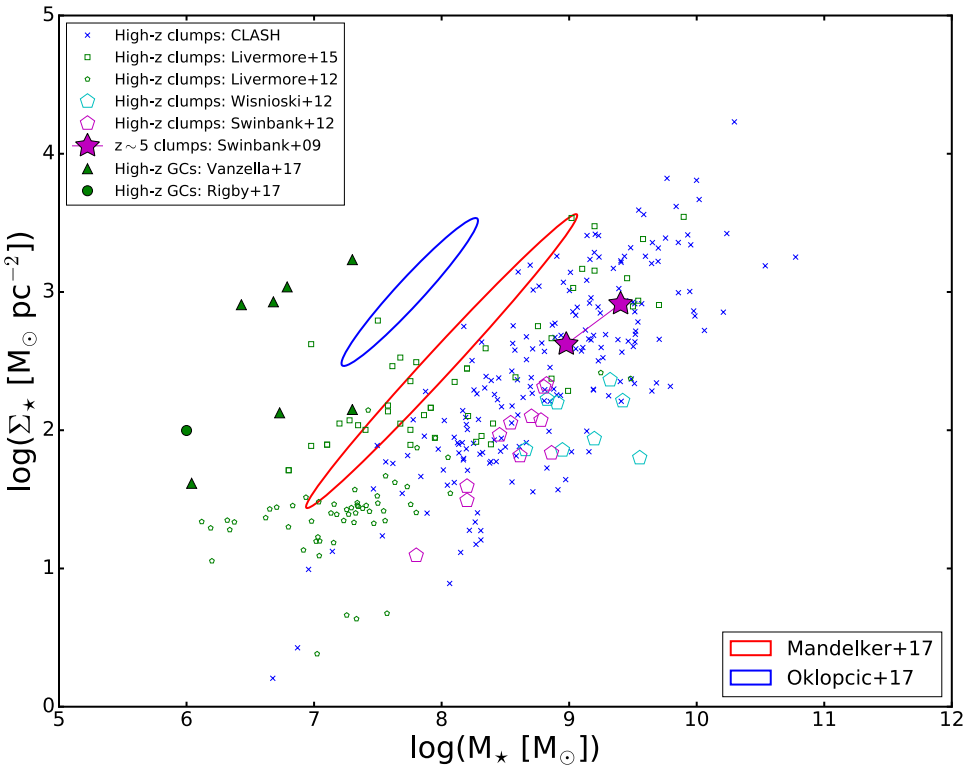


Figure 5.12: The observed and the simulated clumps in the fundamental plane. The blue and red ellipses denotes the distribution of the simulated clumps from Oklopcic et al. (2017) and Mandelker et al. (2017).

performing a detailed investigation of effects of stellar feedback.

5.5.3 Comparison of Radial Variation of Clumps

If the clumps are long-lived, they would continue accreting gas in the course of migration to grow their mass up to a factor of 3 (Mandelker et al., 2014). The sSFR of clumps that reach the inner region of disc should be lower than those newly-born clumps in the outer region. Thus radial gradients in stellar mass

and sSFR with logarithmic slope -0.65 ± 0.13 and $+0.68\pm0.04$ are expected from simulations (Mandelker et al., 2014). From Figure 5.4, we do not observe any significant trends in stellar mass, color and sSFR at $z_s \sim 2$, which is partilly due to the insufficient sampling of massive galaxies with $\mathrm{M}_\star > 10^{10}\mathrm{M}_\odot$ thus the insufficient sampling of clumps in the innermost region. The bias brought in by the geometric effect also prevents us from performing robust comparison. So we are not able to draw conclusions on this comparison on radial variation. Future clump samples that include much larger numbers of clumps from massive galaxies are needed to assess the consistency with simulations based on the radial variation of physical properties.

On the other hand, we compare the radial distribution of sSFR and surface mass density $\Sigma_\star$ at $z_s \sim 2$ with another observation from Guo et al. (2012) which computes H band flux-weighted center as their galactocenter, and the predicted radial distributions for short-lived clumps from Oklopcic et al. (2017) in Figure 5.13. For radial distribution of sSFR, three data sets are very consistent within uncertainties; while our radial distribution of $\Sigma_\star$ is consistent with that of Guo et al. (2012). The predicted $\Sigma_\star$ by short-lived clump scenario in the innermost region is noticeably higher than that of the observations, which has already been seen from the comparison in the fundamental plane. However, the gradient or slope in the innermost region of three data sets appears to be still consistent. Therefore, as argued in Oklopcic et al. (2017), the radial vari-

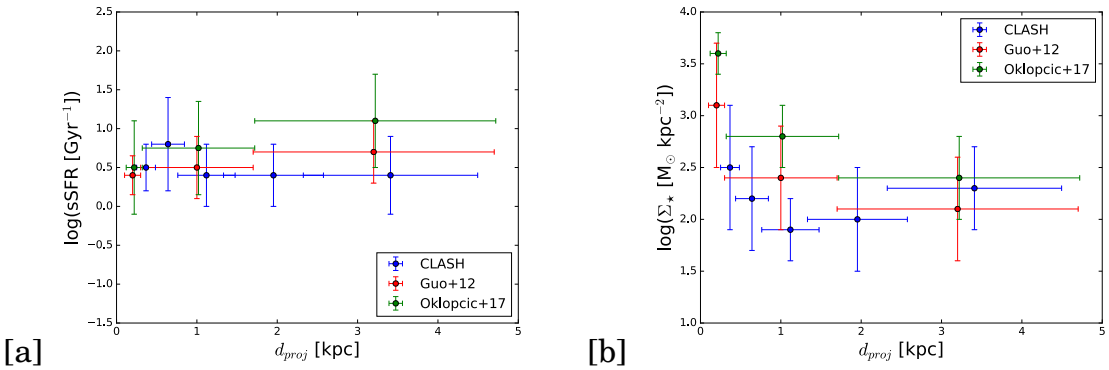


Figure 5.13: The comparison between the observed and simulated radial distribution of sSFR and surface mass density of clumps.

ation of the physical properties of clumps appears to be less informative than expected in helping to disentangle different model predictions on the lifetime of clumps. More informative observables that can efficiently distinguish the simulated results from the long-lived and short-lived scenario are urgently needed. Oklopcic et al. (2017) suggests that gas fraction predicted by short-lived scenario is $\sim$ 1 dex higher than that from long-lived scenario at all radius, thus could be an observable to probe. Unfortunately, we are not able to measure the gas fraction in this study and future spectroscopic data are needed to address this issue.

5.6 Discussion

The clump migration scenario was motivated by the VDI model and proposed as a promising mechanism for the mass assembly of galactic spheroidal.

To justify the feasibility of the scenario, two main issues must be addressed:

1. Can clump survive stellar feedback?

2. If it is long-lived, will the clump migration occur?

In this section, based on the results described in previous sections, we will discuss these two issues in detail.

5.6.1 Is the clump long-lived or short-lived?

The most convincing evidence of the clump lifetime would be by definition – the stellar age of the clump. However, the photometry only based age determination suffer much from the degeneracy between age, metallicity and dust: a red color can be recovered by either of the three. The uncertainty in the age from SED fitting is consequently very large. Even if the degeneracy is broken and the measurement is reliable, the starlight from young stellar populations in clumps due to the continuous accretion of fresh gas and star formation during the process of migration would bias the age estimation young as well. The uniqueness of our study is that we are able to measure the evolution of the radial distribution spanning large redshift ranges. So we could infer the lifetime of clumps from the evolution of other physical quantities.

The short-lived clumps predicted by simulations can only survive for $\sim$ 10-50 Myr. Such short life time would lead to some unique observational features:

1. The short life time only allows clumps to migrate for a short distance and quickly be disrupted by the stellar feedback and disperse into the inter-clump medium. Therefore, a short-lived clump can not exist in the inner region of the galactic disk[1]. Mandelker et al. (2017) has investigated the radial distribution of the long-lived and short-lived clumps during the migration. They find that few short-lived clumps can pass the $d_{crit} = 0.5R_d$ where R_d is the radius of the disk. Therefore, the occurrence of the clumps within $d_{crit} <= 0.5R_d$ can be used to test the lifetime of the clumps.

2. Large dispersion in physical properties is not favored by short-lived clumps. e.g. the short-lived clumps will always look blue and have high sSFR due to the transient nature. They also do not have sufficient growth in their mass or surface mass density.

3. Unless the stellar populations in short-lived clumps differ dramatically between high-z and low-z, there should not be evolution in average color and sSFR of short-lived clumps.

Correspondingly, the features of the predicted long-lived clumps are:

1. Long-lived clumps should cover a larger area of the galactic disk than the short-lived clumps.

[1]This argument is based on an assumption that most of the clumps were born in a star-forming ring in the outer region of disc and migrate toward the galactocenter, which is valid in high-z massive star-forming galaxies. It is not clear if it also applies in smaller galaxies.

2. Long-lived clumps should exhibit a larger dispersion in their physical properties than that for short-lived clumps.

3. The color and sSFR of long-lived clumps should respectively become redder and lower as redshift decreases.

In Figure 5.4,5.5,5.6,5.7, we can clearly see that at $z_s > 2.5$, at least half of the clumps cross the $d_{crit} = 0.5R_d$ line; whereas at lower redshift, there are also a few clumps that cross the line. We identify the increasing dispersion in color and sSFR as discussed in Section 5.4.2 which poses another evidence to support the long-lived scenario. We do not observe significant overall shifts in color and sSFR, but since the clumps are selected from rest-frame UV instead of rest-frame optical, there might be some redder clumps with less star-forming activity escape from our detection. The mass distribution and properties in the fundamental plane of observed clumps are also more consistent with that of the simulated long-lived clumps as discussed in Section 5.5.1 and 5.5.2. Moreover, the clumpy fraction of galaxies is related to the duty cycle of clumps. The probability of observing clumpy structure from a randomly selected galaxy is $f = t_{duty}/t_{obs}$, where t_{obs} is the observability timescale. Our lensing selected sample is by no mean complete. However, we have identified clumps from 98 out of 106 galaxies. Though the high occurrence could also be interpreted as a rapid replenishment cycle of short-lived clumps, that interpretation appears not supported by the increasing dispersion in sSFR and color as redshift de-

creases. Thus the high occurrence would appear to support a long duty cycle, thus the long-lived scenario of clumps.

5.6.2 Does the Clump Migrate Towards the Galactic Center?

As we discussed in Section 5.4.1, the most indisputable evidence of clump migration is the radial evolution of clumps, which we have not detected. Therefore we conclude our data are consistent with the hypothesis that the clumps in less massive galaxies ($\mathrm{M}_\star < 10^{10}\mathrm{M}_\odot$) do not migrate or migrate very slowly. Interestingly, we note that all the single clumps at $z_s \sim 3$ in Figure 5.8 are very massive, whose mass is approaching the typical mass of local bulges. The results of stellar mass function (SMF) in Section 4.9.3 have revealed that the high mass end of clump SMF is consistent with the SMF of local bulges of intermediate/late-type galaxies, and the consistency between the derived CLASH scaling relation and those from CANDELS and simulations on host galaxies of clumps also suggests that the massive CLASH clumps at high-z and the core of SFGs at high-z are likely to be the same thing. In another words, we are likely witnessing the main stage of the mass assembly of the galactic bulges $z_s > 3$, which are observed as massive clumps in SFGs at high-z.

If the clumps are long-lived and not migrate, rather than being quickly dis-

rupted and dispersing into the inter-clump median or merging into a larger systems, it implies that the main way of assembling mass for the massive clumps could only be continuous gas accretion and star formation, if the gas reservoir can be replenished quickly via cold stream. Interestingly, our studies of UV luminosity function and stellar mass function of clumps have shown that the most massive clumps are also the rest-frame UV brightest. If these massive clumps are indeed the progenitors of the local bulges, our data appears to support that the bulges form via "wet" or dissipative process instead of "dry" growth in terms of the clump coalesce. Such dissipative process is consistent with the hypothesis of "compaction" process (Dekel et al., 2009b) which used to explain the origin of the "blue nuggets" and "red nuggets" galaxies. The dissipative growth of clumps is also consistent with our proposed evolutionary track proposed in our fundamental plane study in Section 4.7.3, in which the clumps assemble their stellar mass along the $\mathrm{R_e} \sim 500\mathrm{pc}$ line and the small scatter along the constant effective radius line is interpreted as gradual growth via gas accretion.

5.7 Summary and Conclusions

In this chapter, we study the radial distribution of clumps and assess its compatibility with the proposed clump migration scenario. We divide the full

clump sample into high and low mass samples by the median stellar mass of their host galaxy, and measure the radial distribution and radial variation of physical properties of clumps for the full sample and subsamples, respectively. We do not identify any significant radial evolution thus our data is consistent with that at least in less massive galaxies the clumps do not migrate or migrate very slowly. We do not detect any statistically significant radial gradient in physical properties of clumps in the high and low mass samples. We find the dispersion of physical properties becomes larger as redshift decreases, which is consistent with the long-lived scenario. We compare the dynamical properties of the observed clumps with the simulated clumps in the fundamental plane: The long-lived simulated clumps are more consistent with the observed clumps, whereas the short-lived clumps are intermediate objects between the observed clumps and star clusters. The geometric effect could potentially bias the radial measurement but we find that our conclusion is not significantly affected by this geometric effect. The null detection of clump migration signal along with the higher values of characteristic mass and UV luminosity at higher redshifts suggest that the central bulges are likely built via some dissipative process, which is more consistent with the gradual, steady growth via gas accretion than the clump coalesce.

Chapter 6

Overall Summary and Conclusions

In this thesis, I have presented four studies that are based on the strong lensing effects of the cluster of galaxies: 1. the lensing efficiency of cluster of galaxies and its dependence on various physical effects; 2. the integrated physical properties of the lensed galaxies and their morphology; 3. the physical properties of substructure in high redshift galaxies and their evolution; 4. the radial distribution and radial variation of physical properties of the substructure and their implications on galaxy evolution.

In Chapter 2, we developed an algorithm to find and characterize gravitationally lensed galaxies (arcs) to perform a comparison of the observed and simulated arc abundance. Observations are from the Cluster Lensing And Su-

pernova survey with Hubble (CLASH). Simulated CLASH images are created using the MOKA package and also clusters selected from the high resolution, hydrodynamical simulations, MUSIC, over the same mass and redshift range as the CLASH sample. We derive a lensing efficiency of 4 ± 1 arcs (with length $\geq$ 6 and length-to-width ratio $\geq$ 7) per cluster for the X-ray selected CLASH sample, 4 ± 1 arcs per cluster for the MOKA simulated sample and 3 ± 1 arcs per cluster for the MUSIC simulated sample. The observed and simulated arc statistics are in full agreement. We measure the photometric redshifts of all detected arcs and find a median redshift z_s = 1.9 with 33% of the detected arcs having $z_s > 3$. I find that the arc abundance does not depend strongly on the source redshift distribution but is sensitive to the mass distribution of the dark matter halos (e.g. the c – M relation). Our results show that consistency between the observed and simulated distributions of lensed arc sizes and axial ratios can be achieved by using cluster lensing simulations that are carefully matched to the selection criteria used in the observations.

In Chapter 3, we study the integrated physical properties and morphology of strongly lensed galaxies at high-z. The study is based on a sample which is composed of 106 lensed galaxies which have been detected by our arc-finder from 20 X-ray selected CLASH clusters. We measure their photometric redshift and extract the photometry after correcting for lensing magnification. We use iSEDfit to model the SEDs of the lensed galaxies to infer their integrated

physical properties such as stellar mass, SFR and sSFR. We compare the distributions of stellar mass and SFR of our sample with the CANDELS UDS survey. We find that our sample preferentially selects galaxies with intermediate mass and mild SFR, with a median stellar mass of $10^{9.4}\mathrm{M}_{\odot}$ and SFR of $4\mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$. We compute the sampling volume as a function of redshift and lensing magnification. We perform the cell-by-cell based photometry extraction and SED fitting, and create the spatially-resolved maps for various physical properties. We study the morphology of the lensed galaxies by exploring their position on Gini-M20 plane. We find most of the CLASH lensed galaxies lie in the irregular region. We find that the morphological difference of galaxies between the rest-frame UV and optical images at $z_s > 1.5$ is smaller than that at $z_s < 1.5$. We construct a full morphological catalog by combining the rest-frame optical morphological parameters at low-z and rest-frame UV morphological parameters at high-z. We compare the parameters at different redshifts and find that at high-z, galaxies generally have irregular and clumpy structure and no significant evolution in morphology has been detected. We find that for galaxies with mass less than $10^{10}M_{\odot}$ at high redshift, the morphology is not much affected by their mass. We suggest that there could be some critical mass above which the potential well of the galaxies starts to significantly affect their morphology.

In Chapter 4, we study the substructures (clumps) in high redshift galaxies. 193 clumps have been detected from rest-frame $2200\mathring{A}$ image of 106 CLASH

lensed galaxies via devised clump-finder. We measure their physical properties such as stellar mass, SFR, sSFR, color and size. We find that the distribution of stellar mass of clumps correlates with that of the host galaxies. The dispersion in the distribution of various physical properties increases as redshift decreases, which implies the diversity in dynamical range and stellar population of clumps increase at lower redshifts. We find that the clumps at higher redshifts are more compact than those at lower redshifts. The evolution in compactness is insensitive to the selection of clumps at different scales. We compare the clumps with other spheroidal stellar systems. The clumps fall onto the galaxies branch on the fundamental planes rather than the branch of star cluster. The dynamical properties of clumps are very similar to those of dwarf spheroidals. The proposed origin of the dwarf spheroidals leads us to further propose that there could also be evolutionary track linking the dwarf spheroidals and high-z clumpy galaxies. We measure the rest-frame 2200$\mathring{A}$ luminosity function and stellar mass function of clumps. The trend in characteristic magnitude and stellar mass of rest-frame UV luminosity function and stellar mass function of clumps correlates well with the trend in stellar mass of the host galaxies. The evolution in faint end slope of luminosity function becomes steeper as redshift increases which follows the trend of star-forming galaxies in high-z; the low mass end slope of stellar mass function remains flat except for $z_s \sim 1$, which implies that the clumps form in a scale independent

way. The stellar mass function of clumps at high-z ($z_s \sim 3, 4$) is consistent with that of local late-type galaxies at high mass end ($\mathrm{M} > 10^{9.5}\mathrm{M}_{\odot}$), whereas the bulge mass function declines more rapidly than that of clumps at low mass end ($\mathrm{M} < 10^{8}\mathrm{M}_{\odot}$), which suggests the possible connection between the most massive high-z clumps and local bulge systems. The multiplicity (number of clumps per galaxies) of clumps is insensitive to the physical properties of the host galaxies, and also not much affected by different selection in dynamical range. The average rest-frame UV contribution of clumps to host galaxies is $\sim$ 10% – 20%; the average stellar mass contribution of clumps to host galaxies is $\sim$ 25% – 40%, which is much larger than the typical value at lower redshift. This value is consistent with well with the identified bulge clumps (central dense concentration) in numerical simulations, which again suggests link between the high-z clumps and local bulge systems.

In Chapter 5, we study the radial distribution of clumps and assess its compatibility with the proposed clump migration scenario. We divide the full clump sample into high and low mass samples by the median stellar mass of their host galaxy, and measure the radial distribution and radial variation of physical properties of clumps for the full sample and subsamples, respectively. We do not identify any significant radial evolution thus our data is consistent with that at least in less massive galaxies the clumps do not migrate or migrate very slowly. We do not detect any statistically significant radial gradient in

physical properties of clumps in the high and low mass samples. We find the dispersion of physical properties becomes larger as redshift decreases, which is consistent with the long-lived scenario. We compare the dynamical properties of the observed clumps with the simulated clumps in the fundamental plane: The long-lived simulated clumps are more consistent with the observed clumps, whereas the short-lived clumps are intermediate objects between the observed clumps and star clusters. The geometric effect could potentially bias the radial measurement but we find that our conclusion is not significantly affected by this geometric effect. The null detection of clump migration signal along with the higher values of characteristic mass and UV luminosity at higher redshifts suggest that the central bulges are likely built via some dissipative process, which is more consistent with the gradual, steady growth via gas accretion than the clump coalesce.

As mentioned in the previous Chapters, the main motivation of my works to investigate the sub-galactic structures (clumps) at high redshift is to answer the following questions:

1. How do galaxies at high redshift with irregular, clumpy morphology transform to today's galaxies with regular morphology? or, how is today's Hubble Sequence settled?

2. Do the sub-galactic structures (clumps) play important role in the mass assembly of galactic bulges?

To the first question, some studies have pointed out that the decrease of fraction of clumpy galaxies might be related to the increase of the circular velocity to velocity dispersion ratio V_c/σ (Guo et al., 2015). However, this only answers part of the question: a galaxy was likely born as less clumpy galaxy at lower redshift than at higher redshift. It does not answer the question: *where do those sub-galactic structures (clumps) we have observed in high redshift galaxies end up at?* Our studies have shown that the clumps are likely to be long-lived, so they do not end up dispersing into the inter-clump medium. The VDI and clump migration models suggest that the clumps in massive galaxies likely end up at central bulges via clump migration. However, our studies have shown that the clumps in less massive galaxies are long-lived and do not migrate or migrate very slowly. Where do the clumps in these galaxies end up at? Our study of clump fundamental plane suggest that the answer could be: the clumps in at least some of the high-z low-mass galaxies likely end up with dwarf spheroidals due to the morphological transformation by the environment effects (ram-pressure stripping, strangulation, tidal harassment, etc), which is similar to the long proposed morphological transformation from dwarf irregulars to dwarf spheroidals (Kormendy, 1985). Therefore, the similar stellar mass and size that high-z clumps and local dwarf spheroidals share might not be a coincidence, but the result of an likely evolutionary track. Future larger clump samples that cover broader dynamical range of host galaxies and larger

redshift range can be coupled with the galaxy environment study to assess this hypothesis.

To the second question, we find that the brightest and most massive clumps at $z_s > 2.5$ are likely the progenitors of the local bulges. Our null detection of signal of clump migration strongly favors that the bulges are formed via some dissipative process such as the proposed "compaction" process, instead of the "dry" growth of clump coalesce. In addition, if the evolutionary track between the high-z massive clumps and local bulges indeed holds, it appears to imply that the bulges have already built up most of their mass since $z_s \sim 3$, which is earlier than the typical era of bulge mass assembly $z_s \sim 1-2$ that is commonly proposed in the frame of clump migration.

Combining the above clues, a new picture of galaxy evolution at high-z appears to emerge: at $z_s \sim 2$ which is the peak of the star formation and galaxy merger activity, the low-mass galaxies with $\log(M_\star) < 10$ that possess clumpy morphology generally have two kinds of destinies: they would either fall into some dense environments to have their gas being stripped and star formation activities being quenched, to end up with smaller dwarf spheroidals; or they would continue accreting cold gas from cosmic cold streams to grow into more massive disk galaxies. Then the more massive host galaxies will exert larger gravitational torque on the clumps to drive them towards the galactocenter quickly, though such "dry" mass assembly might not contribute much to the

overall mass budget of the bulges. In both cases, the morphology of galaxies transforms from irregular to regular, which naturally explains the settlement of today's Hubble sequence. Unlike the classic picture of the formation of Hubble sequence, in which the morphological transformation has been accomplished by galaxy mergers, our picture does not require frequent galaxy mergers in forming the galactic bulges and Hubble sequence, which is consistent with the low galaxy merger rate observed at $z_s < 1$. If this picture is correct, it might be useful to explain why there are many bulgeless galaxies (some of them are even very massive disk galaxies) at low redshifts which appears to contradict with the bottom-up hierarchy of galaxy formation. Moreover, it also has interesting implications for the formation of the red-sequence of local galaxies: if majority of low-mass galaxies at high-z assemble their mass via cold gas accretion instead of galaxy merger, the transition between the blue cloud and the red sequence might be a more smooth or even bi-directional process (Kannappan et al., 2009): a red galaxy can accrete gas to form a more extended disk to become blue; whereas in massive galaxies, the infalling gas can be shock heated to quench the star formation activity in the disc to deliver the galaxies into the red sequence (Birnboim & Dekel, 2003; Keres et al., 2005; Dekel & Birnboim, 2006). Future larger and deeper galaxy survey would be valuable to test the validity of this picture.

Appendix A

Integrated Quantized Intensity-Difference Criterion

The following appendices provide further details about the arc-finder algorithm. Specifically, we provide short summaries of the key steps performed to go from the initial science image to the final arc catalog. We begin by convolving our HST images with a square Top-hat kernel with an edge dimension of $0.065''$ to modestly enhance the contrast of the faint and thin arcs. Most source detection algorithms work in intensity space, which means the performance of these algorithms largely depends on how the detection threshold is chosen. A higher threshold will yield a catalog with lower completeness for faint objects while a lower threshold will lead to less precise segmentation and a higher false positive rate. To avoid the non-trivial determination of an optimal detec-

tion threshold, we focus on three very general properties of giant arcs:

1. Giant arcs, like all real astronomical sources, have a net positive amount of flux on average after subtracting off a suitable background level.

2. Giant arcs have substantial angular lengths.

3. Giant arcs are highly elongated objects.

The above general properties imply that, on average, the intensity difference between the pixels belonging to the arc should be positive and the elongated and distorted morphologies of arcs should also be reflected in the angular distribution of these intensity differences. Use of the non-parametric intensity differences has a genuine advantage in the arc detection game: we can, in principle, detect faint structures almost as easily as bright structures. For this key reason, we perform the primary arc detection process in intensity-difference space. To do this, we first lay down a grid of points on the smoothed image, at spatial scale n, that is somewhat larger than the arc widths we wish to find. At each grid point we then determine whether each of its 8 adjacent grid points (up, down, left, right, upper-left, upper-right, lower-left, lower-right) is brighter or fainter than this pixel. We quantify this local set of flux differences by assigning a value of $+1$ for positive difference (the central pixel at grid position (i, j) is brighter than a given surrounding pixel) and a value of -1 for a negative difference (the central pixel at grid position (i, j) is fainter than a

given surrounding pixel). We sum up these values for all 8 directions. A grid point that was brighter than all of its surrounding grid points would thus have a final value of $+8$. A grid point that was brighter than 6 of its surrounding grid points would have a final value of $6 - 2 = +4$. And so on. As arcs are highly elongated, pixels lying along the ridge line of an elongated arc will tend to have at least 4 or 5 adjacent pixels that are fainter than those at a given grid position. The value assigned to these pixels will thus be at least 2 or higher ($5 - 3 = 2$). In general, the brighter pixels in an arc will tend to have higher integrated quantized intensity-difference values than the fainter pixels. Given that some giant arcs may have complex intensity profiles we set the threshold for the integrated quantized intensity difference to be the lowest positive value, which is $+2$. If we adopt a higher positive threshold, we find that some complex arcs are segmented into several smaller arc detections. The threshold of $+2$ is the most conservative in maintaining the overall structural shape of the arc candidates. We note that the exact choice of threshold value, however, does not significantly impact the contents of the final sample of large ($l \geq 6$") and highly elongated ($l/w \geq 7$) arc candidates. The effect of the quantized intensity difference threshold is primarily on the number of small and less elongated sources in the initial detection process.

Choosing a proper grid spacing scale, n, is important. Generally, the spacing scale n should be larger than the typical arc's width, and it should neither be

too large nor too small, to avoid extending the grid points to nearby bright structures or limiting the grid points around the arc rigid lines. To determine the scale, we visually select 58 giant arcs from our CLASH F814W images, and manually measure the arcs' full width at half maximum (FWHM) in the direction perpendicular to their ridge lines [1]. Figure A.1 shows the distribution of the pre-selected arc's FWHM. Note that the median value of these 58 arcs is $0.33''$ and most of the arcs widths are less than $0.72''$. In principle we should traverse as many grid scales as possible to optimize the detection of the arcs, which is computationally expensive. We adopt two different scales: $0.39''$ and $0.78''$, to make sure that both narrow and wider arcs can be effectively detected in a relatively short computational time. The results based on each scale are combined as the input to the next step.

As noise pixels may have regions with zero-valued or negative integrated quantized intensity-difference [2], another obvious advantage is that we are able to effectively clip out noise pixels and make the arc detection task significantly easier, even in the presence of a bright diffuse background, as might be encountered in the halo of a bright foreground cluster galaxy.

[1] To measure the FWHM, we first draw a line crossing the intensity maxima which is perpendicular to the arc's ridge line, then use Gaussian profile to fit the intensity of pixels that fall on the line. We approximate the Gaussian FWHM as the FWHM of the arc.

[2] For noise pixels, if their distributions are independent, the integrated quantized intensity-difference should be equal to 0.

APPENDIX A. INTEGRATED QUANTIZED INTENSITY-DIFFERENCE CRITERION

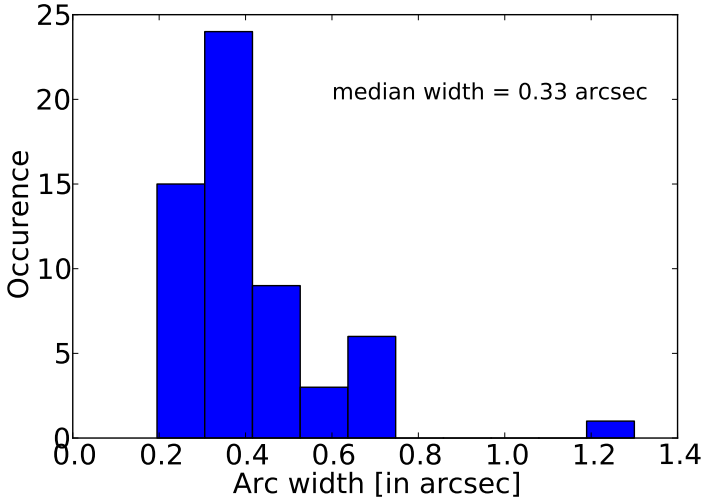


Figure A.1: The FWHM distribution of the pre-selected 58 giant arcs from CLASH F814W images. The median FWHM is around $0.33''$, and most of arc widths are less than $0.72''$. The exception amongst this sample is from the cluster MACS1206, which includes a giant arc with width $\sim 1.3''$.

Appendix B

The Local Intensity Difference Criterion

In certain regions (especially in the inner cores of bright galaxies), applying the integrated quantized intensity-difference criteria only will leave the segments with the diffraction pattern (see Figure B.1 [a]). To suppress these effects, we apply another criterion by comparing the intensity of the central pixel with the mean value of all 8 adjacent pixels over the image. The selected pixels should satisfy two criteria below:

$$\sum_{l,m\in(-n,0,n)} SIGN(I(i,j) - I(i+l,j+m)) \geq 2 \tag{B.1}$$

$$\sum_{l,m\in(-n,0,n)} (I(i,j) - I(i+l,j+m)) > 0 \tag{B.2}$$

APPENDIX B. THE LOCAL INTENSITY DIFFERENCE CRITERION

Figure B.1 [b], [c] show the integrated quantized intensity-difference maps of MACS0717 before and after applying the above criteria. We can see that number of diffraction artifacts in the image is significantly reduced.

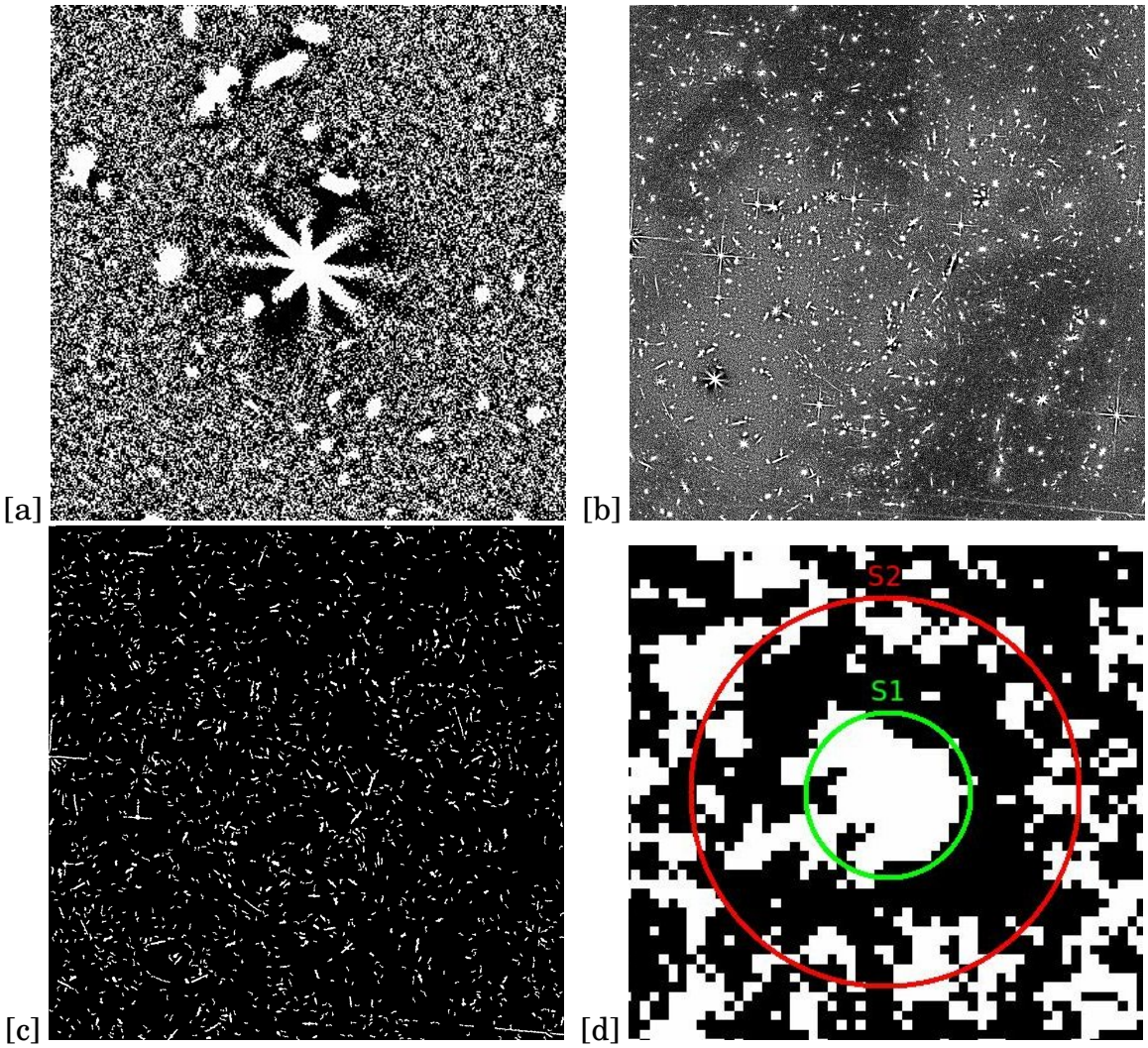


Figure B.1: (a) shows the segment with the diffraction pattern which needs to be suppressed by the local average criteria; (b), (c) are the integrated quantized intensity-difference map of MACS0717 F814W image, before and after applying the local average criteria; (d) illustrates the idea of local selection of contiguous regions: to draw circles with different size on each pixel and calculate the average number density within the circles, and select those grid points which have higher average number density within smaller circles.

Appendix C

Initial Image Segmentation

To identify specific arcs, we need to locate regions of contiguous grid points in the integrated quantized intensity-difference map with sums in excess of $+2$. We have now replaced the challenge of finding objects in intensity space with the task of finding contiguous regions in this quantized intensity-difference space. We avoid using any global selection criterion on number density since the number density varies largely across the whole image. So the contiguous regions are selected by their local number density of the grid points in the quantized intensity-difference space. Based on the simple fact as Figure B.1 [d] shows: if the contiguous region is enclosed by circle S1, the local averaged number density inside S1 must be larger than that inside circle S2 which has the same center with S1 but larger radius. The details of contiguous regions selection are as follows: (1) we make three convolved images using three spher-

ical uniform kernels (k1, k2 and k3) with increasing size ($0.52''$, $1.04''$ and $1.56''$); (2) we subtract an image convolved with a broader kernel from one convolved with a narrower kernel, to obtain two residual images (k2 - k1, k3 - k2); (3) we then select all the pixels which have positive values in both residual images.

The selected contiguous regions include a few small and less elongated blobs that are not real sources. We set an area threshold $A > 100$ pixels and an eccentricity [1] threshold $e > 0.85$ to remove these artifacts. As shown in Figure B.1 [b]. the noise has been suppressed and most giant arcs have been retained.

[1]The eccentricity here is equal to the eccentricity of the ellipse that has the same second-moments as the measured object

Appendix D

Suppression of Diffraction Spikes

Diffraction spikes from bright stars are the features likely to account for most of the false positive detections. The normal way to remove the star spikes is to locate the position of bright stars and then manually mask out the diffraction pattern. Here we adopt a different approach which eliminates the need to know the position of the bright stars or the direction of the spikes in advance.

Our approach is to merge the diffraction spikes with each associated star and eliminate the combined source as a whole. To do this, we enhance the strong intensity gradients near bright stars and their diffraction spikes by applying unsharp masking. The unsharp masking enhances the peak and dampens the wings of the intensity distribution. As a result, some dark halos can

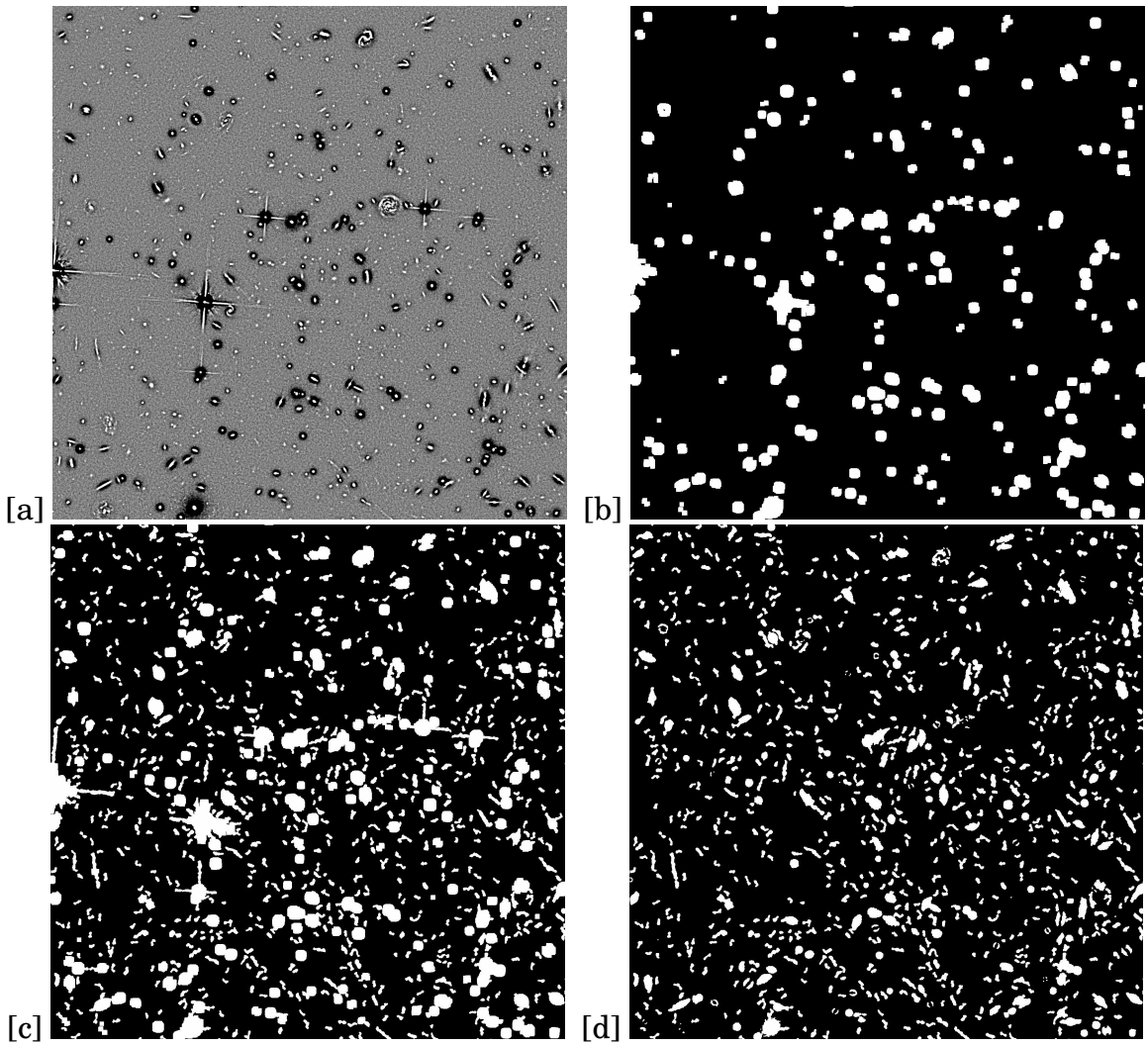


Figure D.1: The black halo regions are identified from the unsharp masked image by setting the threshold to $-0.01e^{-}s^{-1}$. The halo segments are "dilated" (expanded) and combined with the normal detection image so that the stars start to merge with the diffraction spikes. Most of the diffraction spikes can be removed by setting the maximum intensity value of the labeled segments less than $10e^{-}s^{-1}$.

be observed around the stars or bright elliptical galaxies, which are shown in Figure D.1 [a]. Empirically, we note that most of the pixels belonging to the dark halo regions in CLASH data tend to have a intensity value lower than $-0.01e^{-}s^{-1}$, and we use this intensity as a threshold to identify these halos. We then dilate the segmentation boundaries around a dark halo in all directions to fill the gaps between the segments (Figure D.1 [b]), and combine the "dilated" images with the initial segmentation image obtained in Appendix C (Figure D.1 [c]). Most of the diffraction spikes merge with the segments of their mother stars as a result of performing this combination. We then label all the connected components [1]

in the combined image and calculate the maximum pixel intensity of each labeled connected component. Stars typically have maximum intensity values greater than $10e^{-}s^{-1}$, while other objects barely have the maximum intensity value larger than $2e^{-}s^{-1}$, therefore we can conservatively set $10e^{-}s^{-1}$ as threshold to remove those bright stars along with the diffraction spikes (Figure D.1 [d]).

[1] Whether a pixel connects to its neighbors or not is characterized by the pixel connectivity. Usually there are two types of connectivity: 4-connected and 8-connected. 4-connected pixels are connected horizontally and vertically, or diagonally; 8-connected pixels are connected horizontally and vertically, AND diagonally. In terms of pixel coordinates, in 4-connected case, every pixel that has the coordinates $(x \pm 1, y)$ or $(x, y \pm 1)$ is connected to the pixel at (x, y); in 8-connected case, in additional to 4-connected pixels, each pixel with coordinates $(x \pm 1, y \pm 1)$ or $(x \pm 1, y \pm 1)$ is connected to the pixel at (x, y). In this paper, all the adjacent 8-connected pixels are considered to belong to the same connected component.

Appendix E

Final Image Segmentation

The initial segmentation boundaries for objects detected in intensity-difference space tend to have systematically larger surface area than the corresponding segmentation boundaries in pixel intensity space. So we refine the initial segmentation map to correct this small effect. We first define, for each detected segment, a "bounding box" that spans the region from the minimum x, y coordinates to the maximum x, y coordinates. We then iteratively clip out pixels with very high (low) intensity within this box until the pixel intensity reaches convergence at $\pm 3\sigma$ around its median value. We then estimate the local background and noise level within the box. Since faint arcs are most likely missed or broken into small arclets at a high detection threshold, we set the threshold for the re-segmentation to be proportional to the object's estimated local signal-to-noise ratio. Hence, objects with low surface brightness will be remapped us-

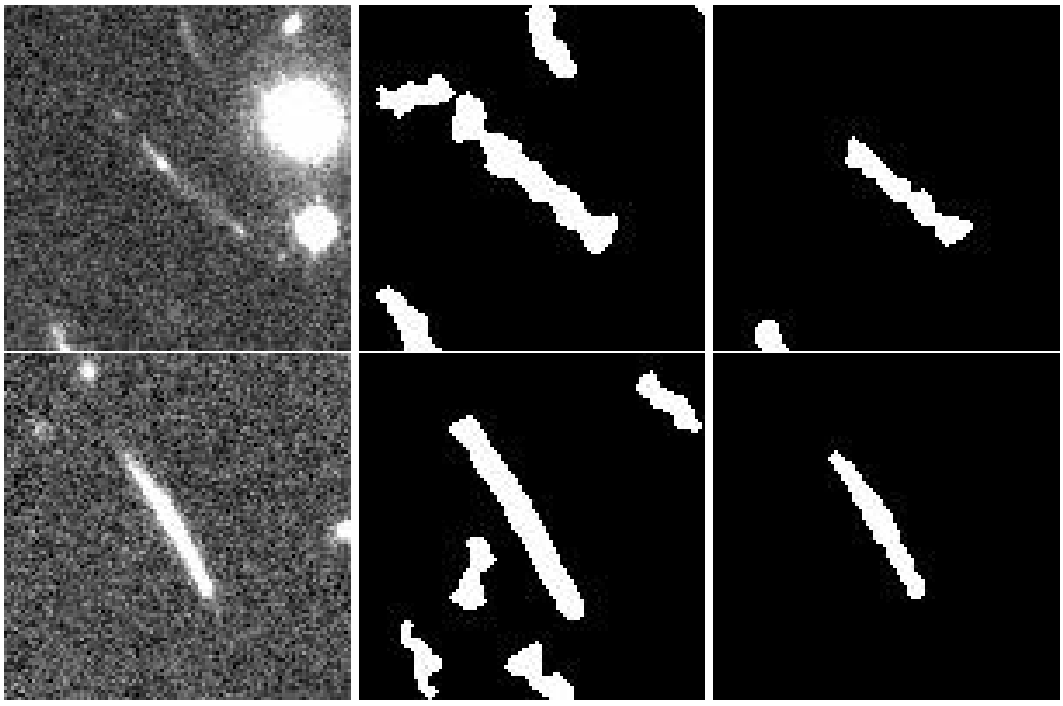


Figure E.1: The $left$ panel shows the original images of arc; the $middle$ panel shows the primary segmentation; The $right$ panel shows the images after the segmentation re-determination. The local background and noise level within the box are estimated and the detection threshold is set to be proportional to its signal-to-noise level.

ing a lower detection threshold than that used for brighter objects, allowing all

sources to achieve their best segmentation (see Figure E.1).

Appendix F

Utilization of the Length, Length-to-Width Ratio and Perimeter-to-Length Ratio

Once all images are processed through the preceding steps, we can begin the arc identification process. We identify giant arcs from among all detected sources primarily by their large ellipticity. For each source, we calculate the total area, perimeter length, and position of the peak intensity from the distribution of all the connected pixels[1]. Using the coordinates of the pixel with the peak intensity value in a given source, we locate the furthest point away from that maximum that is still within the boundaries of the source. We also

[1]We utilize `ndimage` (a Python image processing module) to quickly calculate the mentioned parameters of the detected objects

locate the furthest point away from that point, then calculate the sum of the distances from these two points to the peak pixel position, and take this distance as the length of the segment. There are many ways to define the width of the segment: the image segments can be fitted by simple geometrical figures such as ellipse, circles, rectangles and rings (Miralda-Escude, 1993; Bartelmann & Weiss, 1994); and therefore the width of the segment is approximated by the minor axis of the ellipse, the radius of the circle, the smaller side of the rectangle, or the width of the ring; Dalal et al. (2004); Horesh et al. (2005); Hennawi et al. (2007) approximated the width by dividing the area by its length; Meneghetti et al. (2008) proposed a more robust way to measure the width, by traversing the width profile of the arc and approximating the arc width as the median value of the profile. In this study, considering the computational efficiency, we adopt the former method: i. e. all the giant arcs are treated as rectangles and width = area / length, to determine the width of the segment in this paper. To test whether this definition of width will introduces bias in the measurement of l/w, we use the approach in Meneghetti et al. (2008) to re-calculate the width of all the detected arcs and compare with those in former definition. Figure F.1 shows the comparison of the ratio of two widths with the newly defined width. The dashed line denotes the median value of the ratio which is about 10% higher than that in our definition. Therefore, our l/w (width) measure may be slightly biasing high (low).

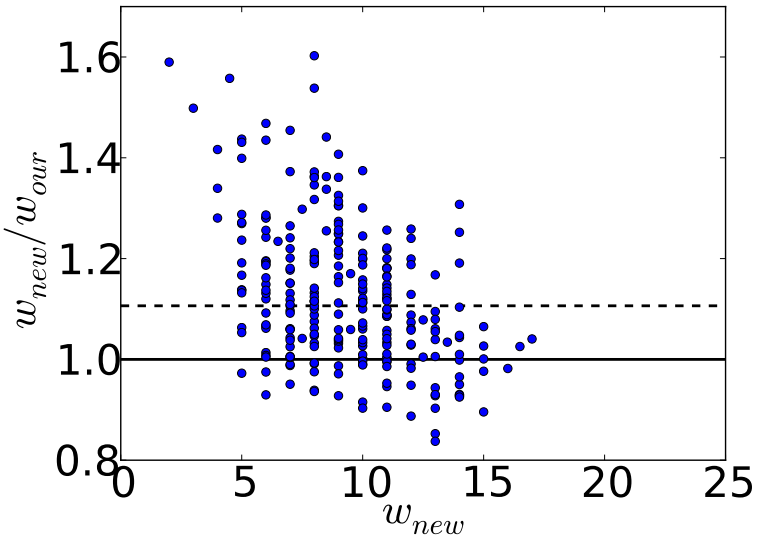


Figure F.1: The comparison of two definition of the width of arc: the y-axis is the ratio of the newly defined width to our width; the x-axis is the newly defined width. The solid line represents the $w_{new} = w_{our}$, while the dashed line denotes the median value of the ratio of the two widths.

The final step is to remove those detected segments that are not very likely to be large lensed galaxies by requiring objects to satisfy three additional criteria [2]: (1) their perimeter-to-length ratio must be ≥ 3; (2) their minimal length must be greater than a fixed value which is discussed in Section 2.3.2; (3) their minimal length-to-width ratio must be greater than a fixed value which is determined in Section 2.3.2. The criterion (1) eliminates elongated objects with irregular morphology and criterion (2) both maintains the consistency with the concept of the "giant" arcs and prevents from the domination of the spurious detection as we will discuss later. We include all objects that satisfy these three

[2]In this study, we do not need to specify the orientation of the giant tangential arcs relative to the cluster center. This allows us to apply our algorithm to less relaxed clusters that may not have a well-defined center.

constraints into our final arc candidate catalog.

Appendix G

The Full Image Catalog of Spatially Resolved Maps in Image and Source Planes

The full image catalog of spatially resolved Maps in image and source Planes for CLASH lensed galaxies with magnification factor > 5 are listed in Figure G.1.

APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES

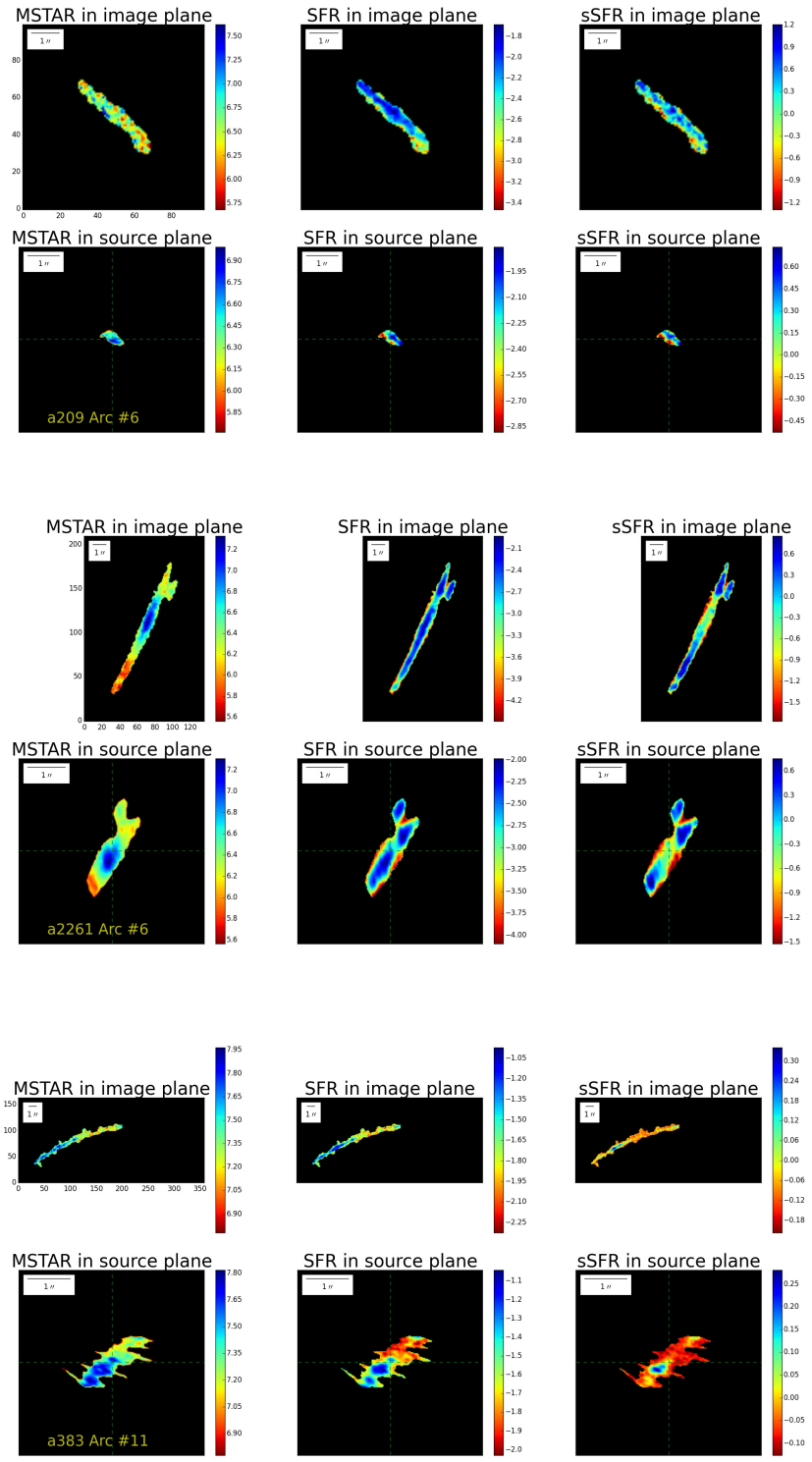
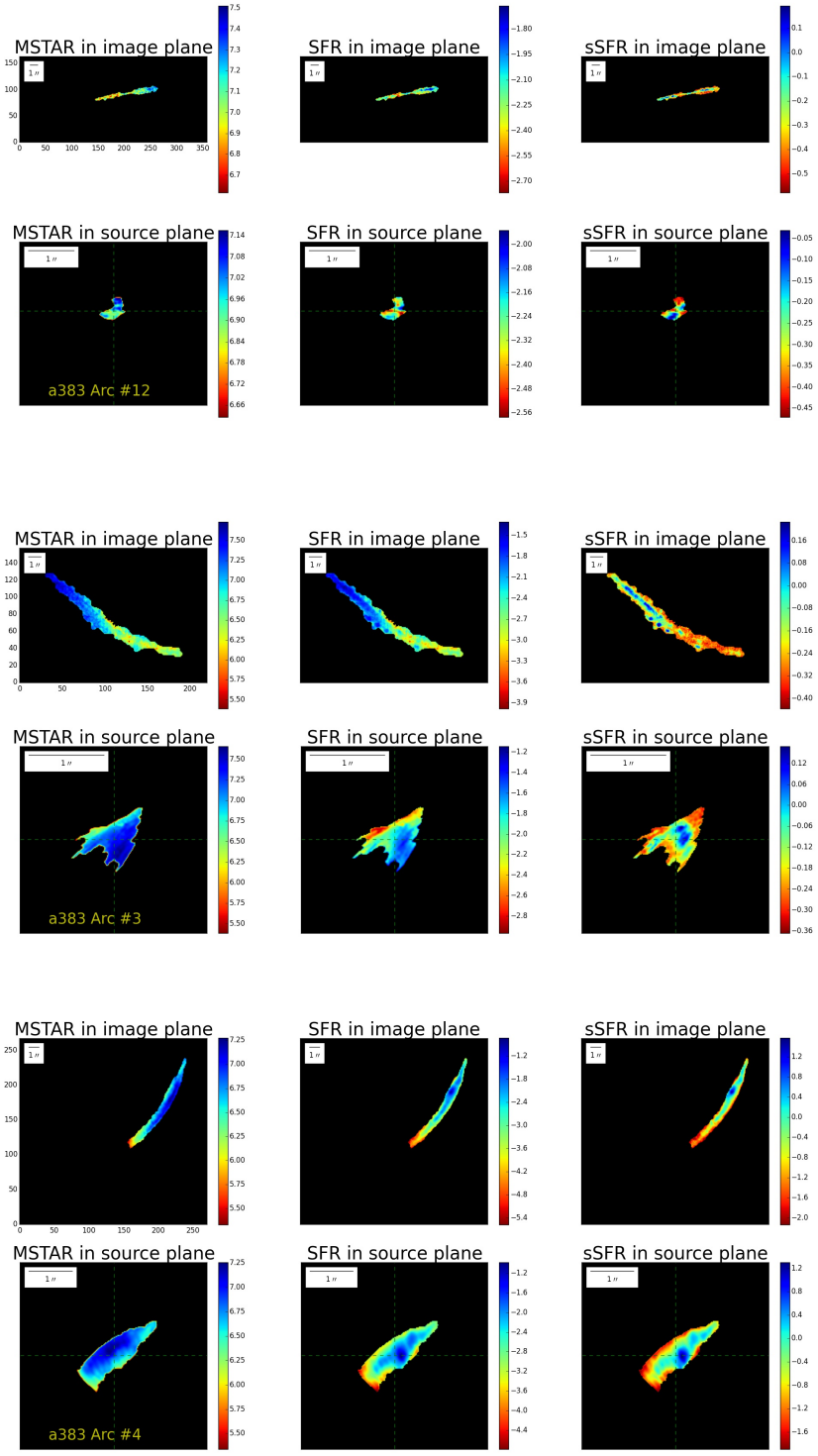
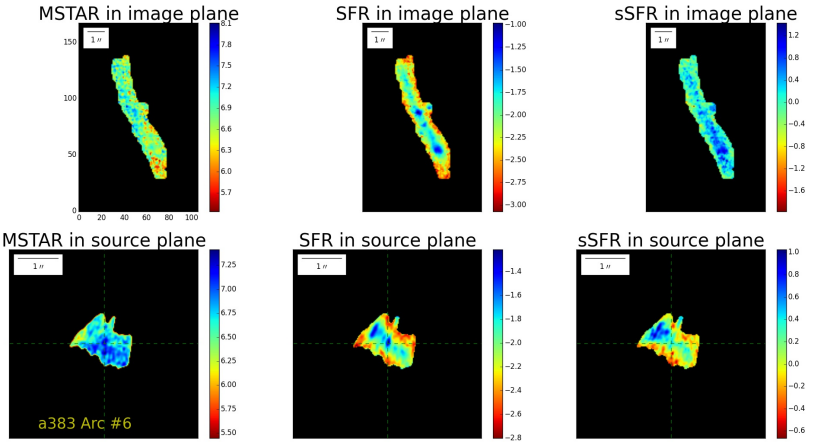


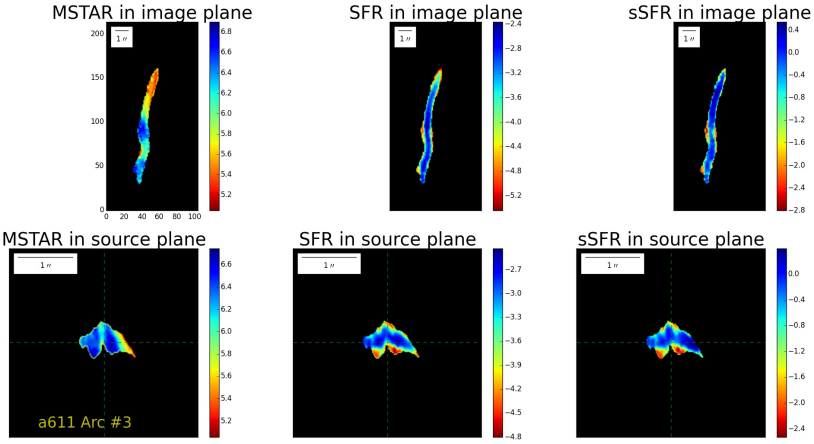
Figure G.1: The spatially-resolved maps of stellar mass, SFRs and sSFR for 53 detected giant arcs with magnification factor > 5. The figures in the upper (lower) panel are the maps in the image (source) plane.

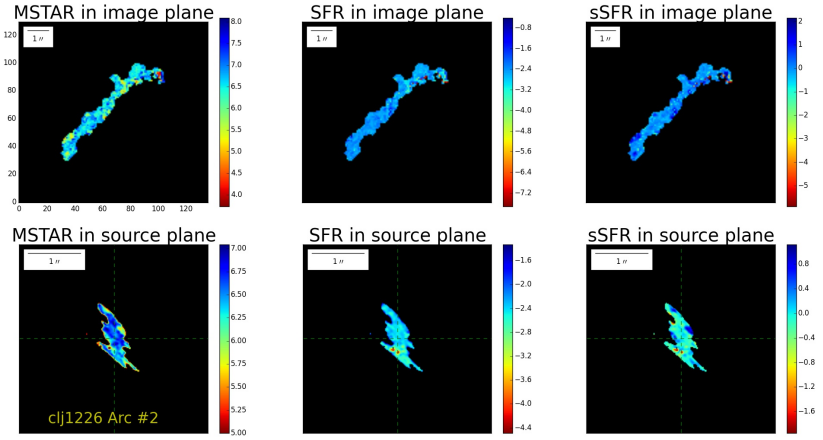
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



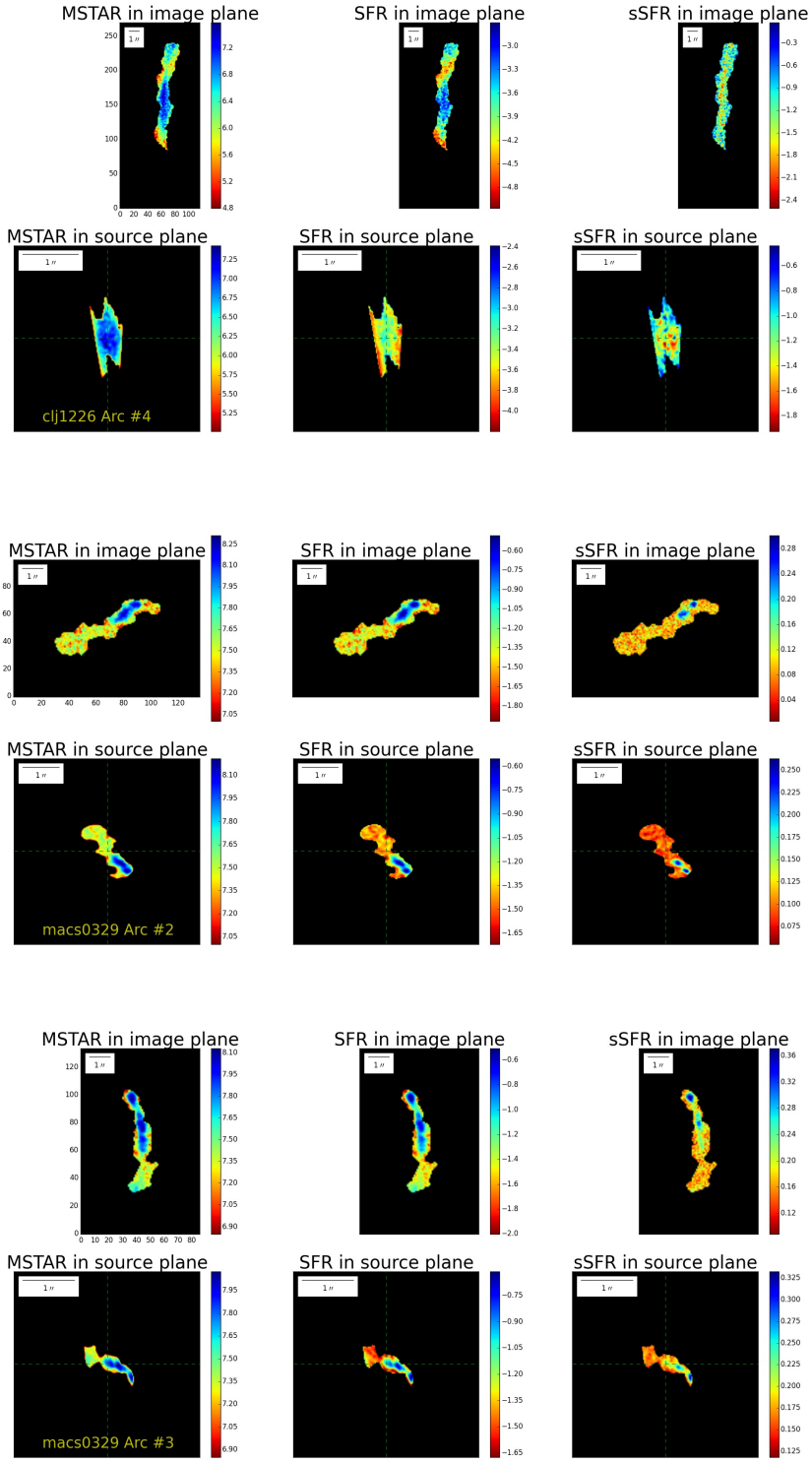
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



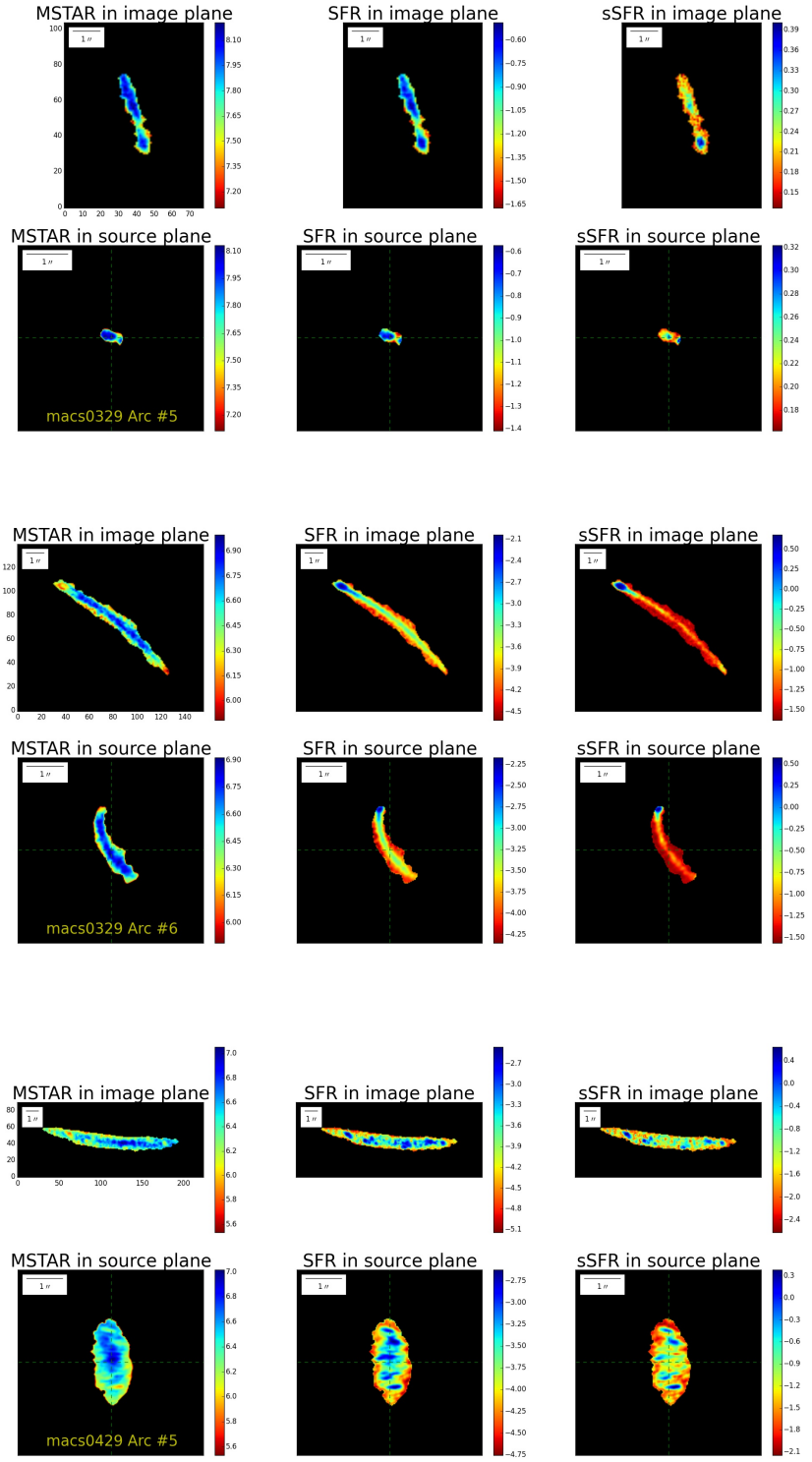




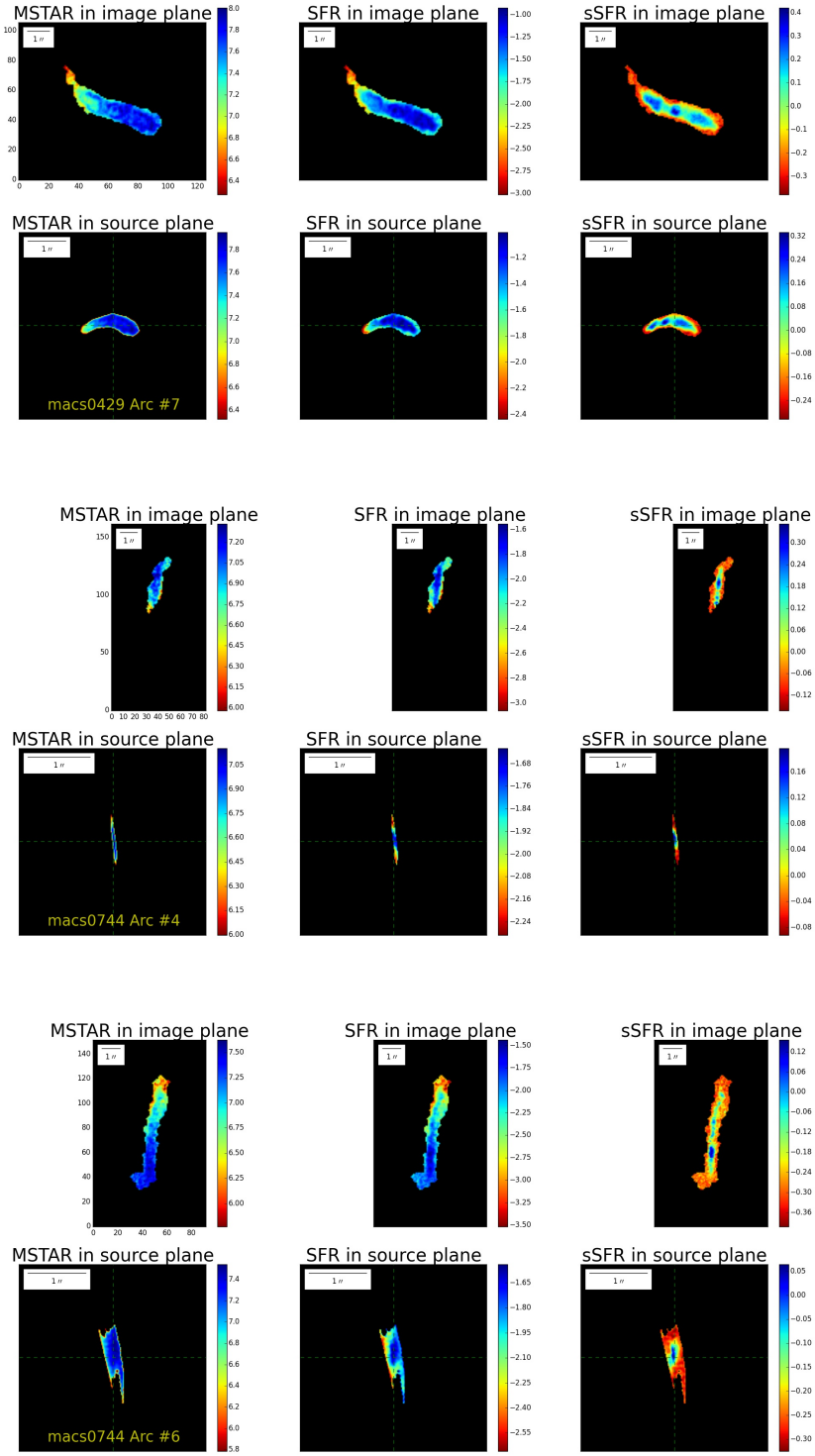
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



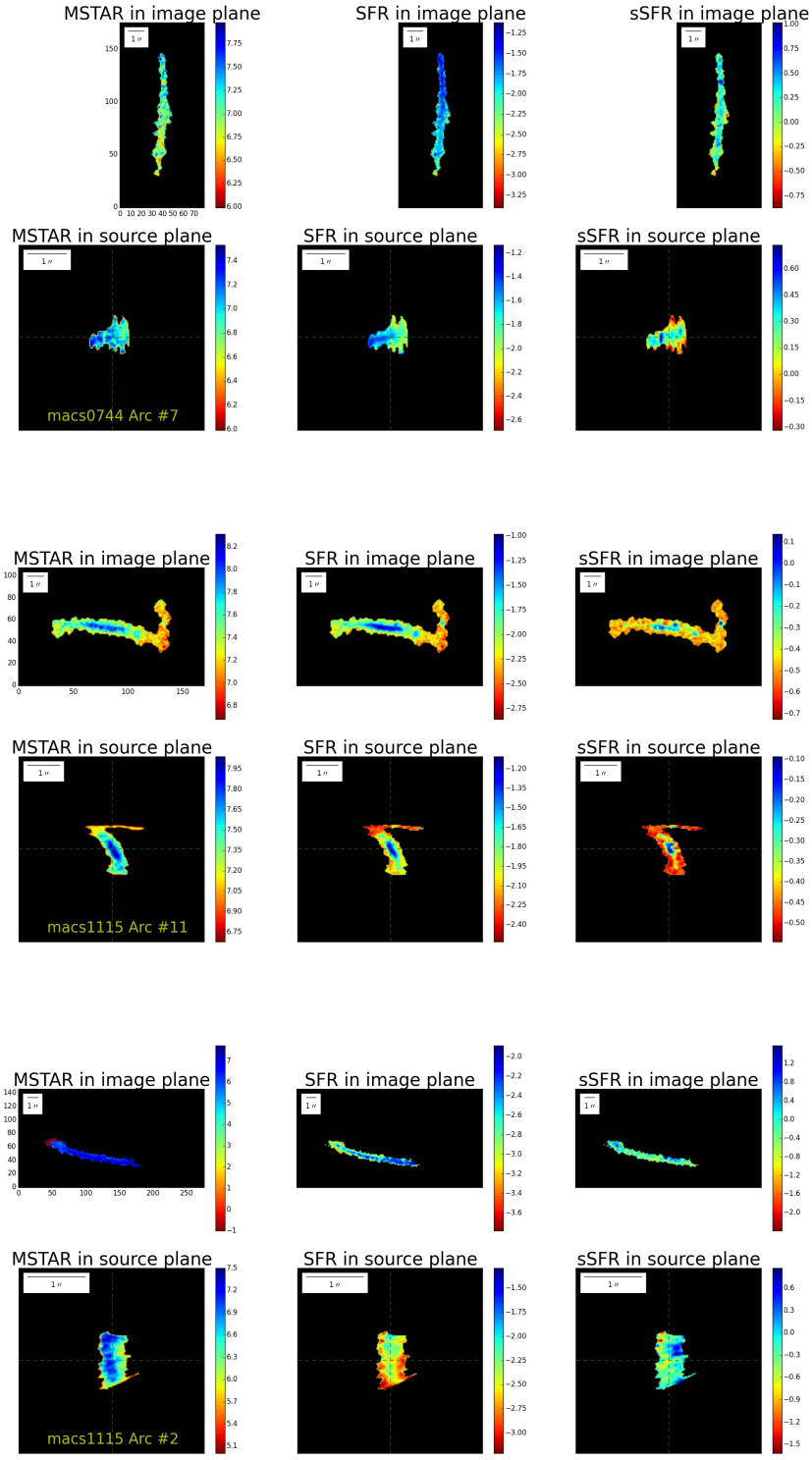
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



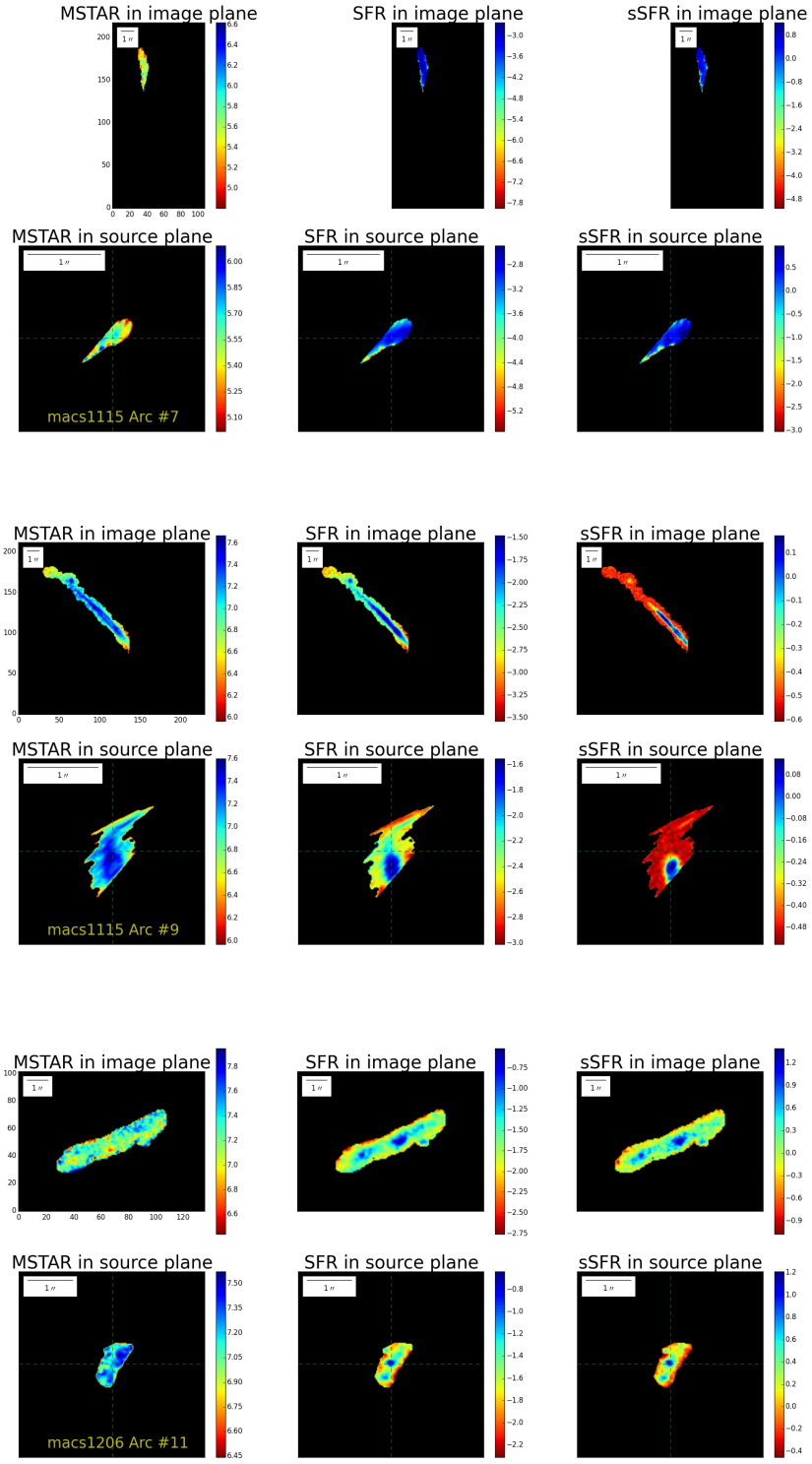
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



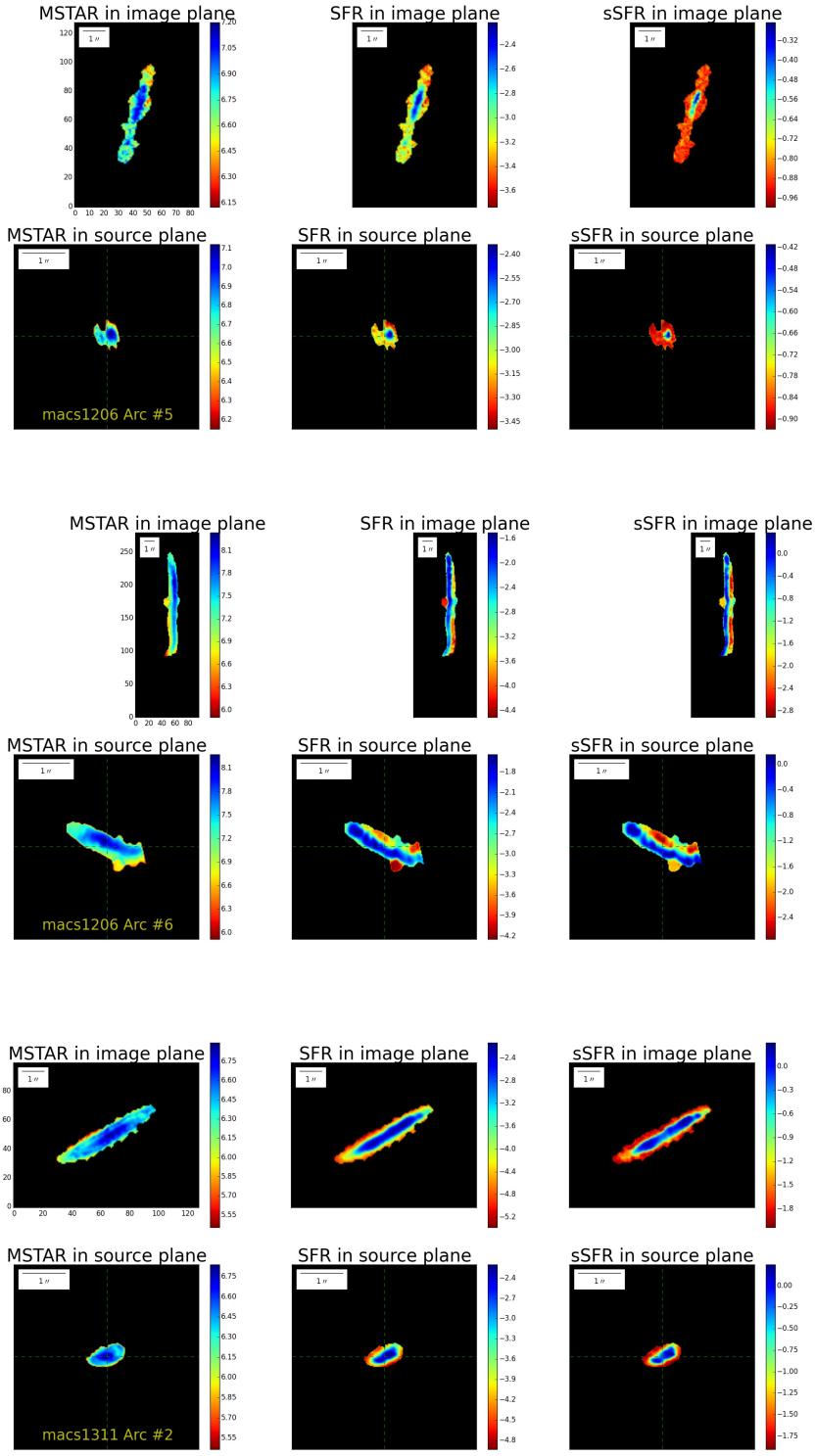
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



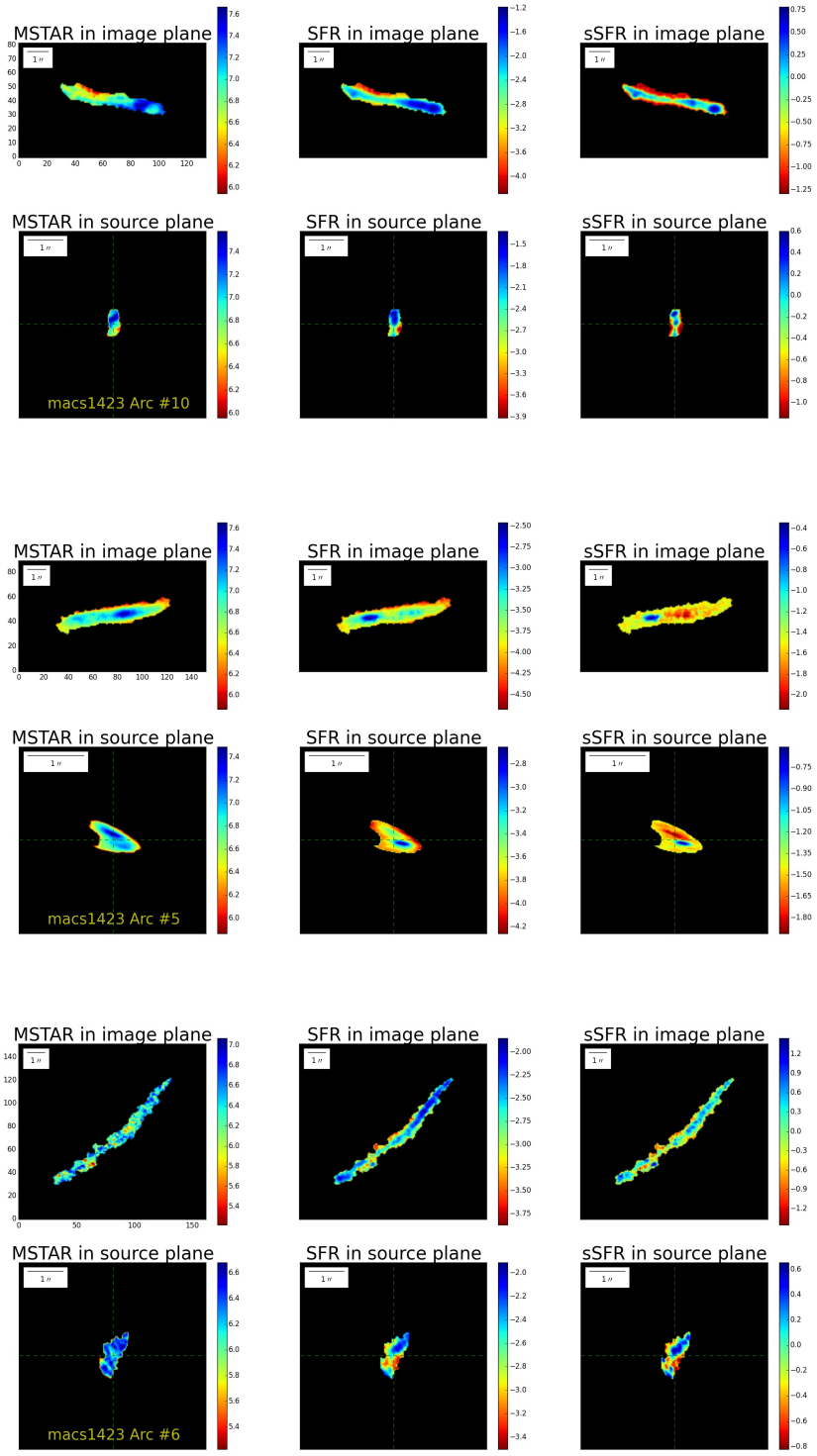
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



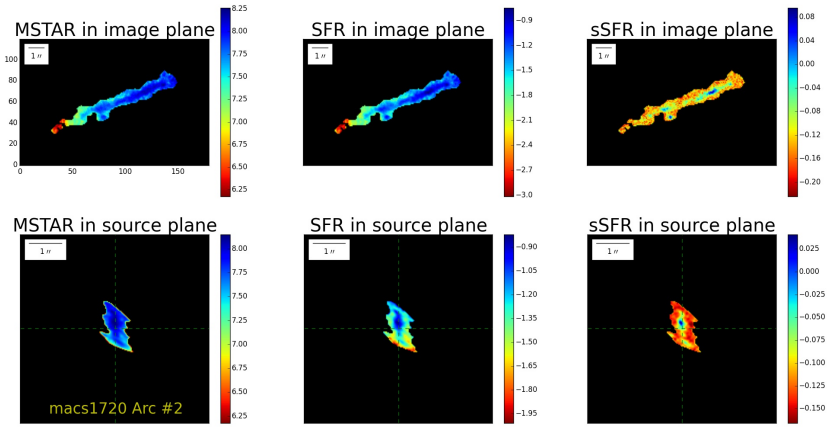
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES

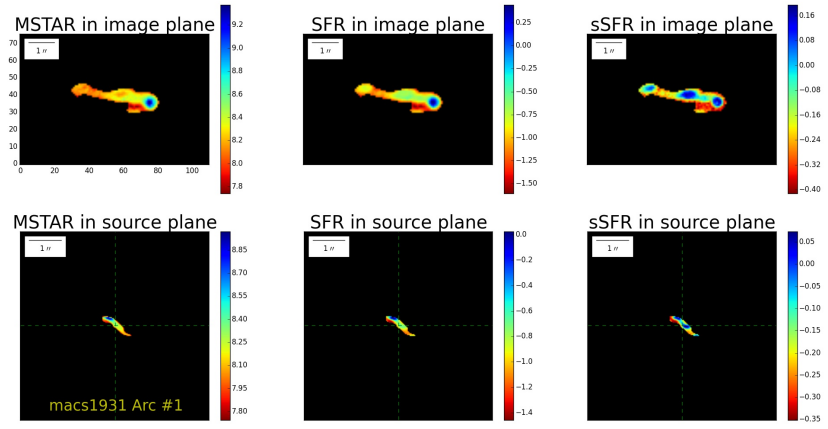


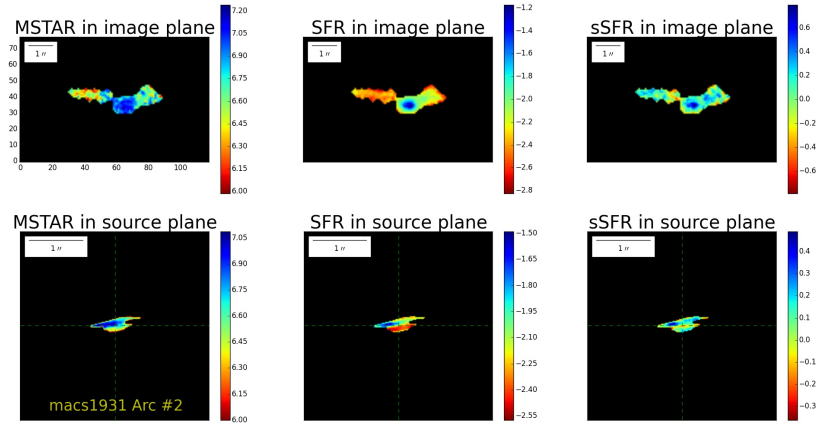
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



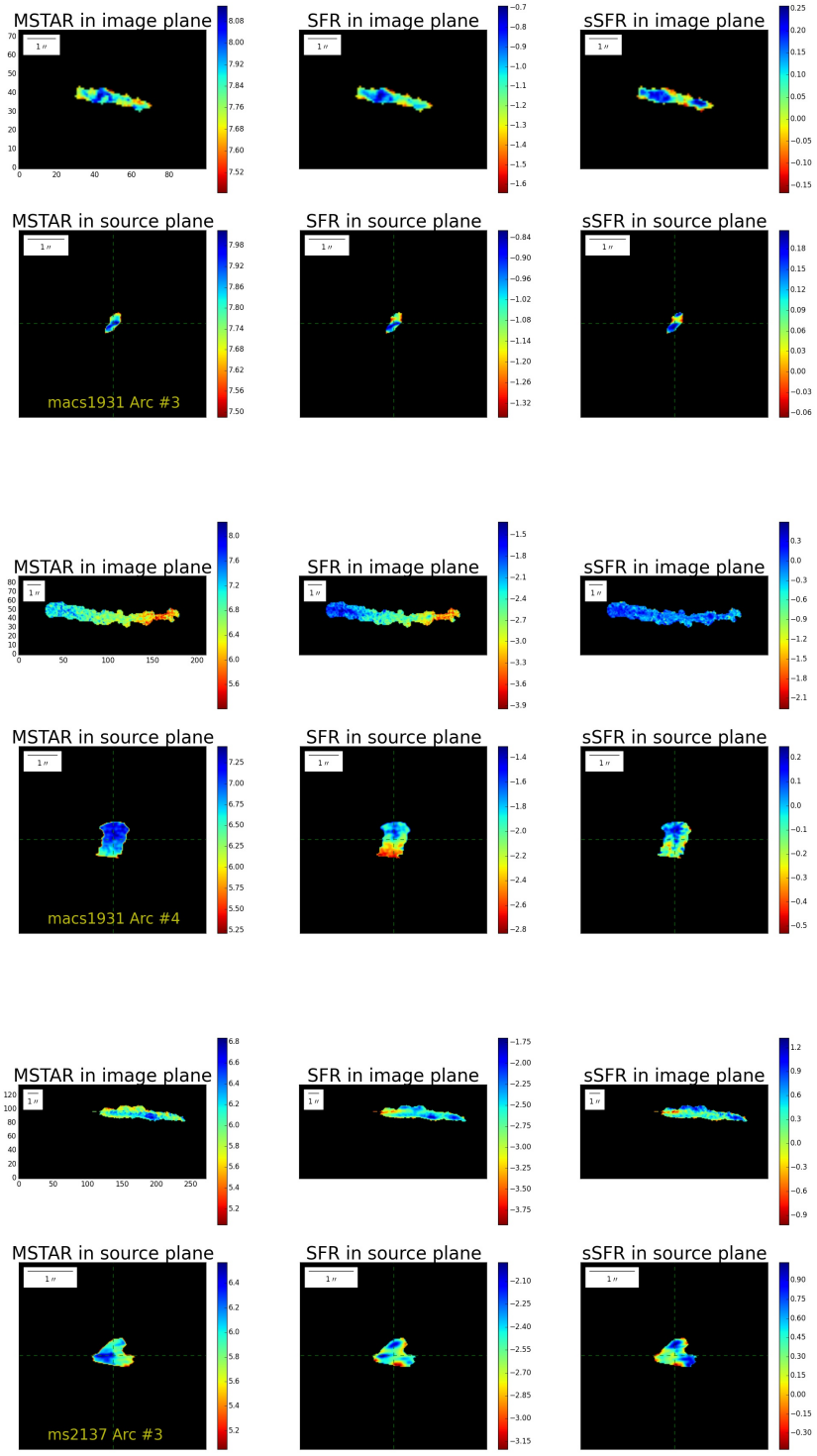
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



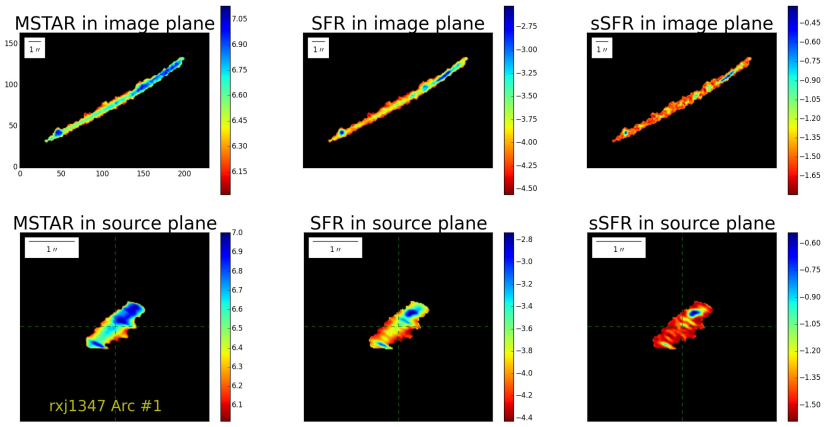


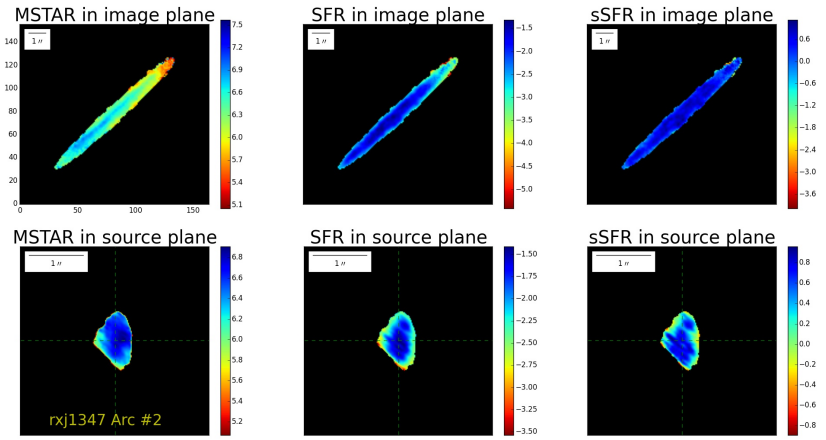


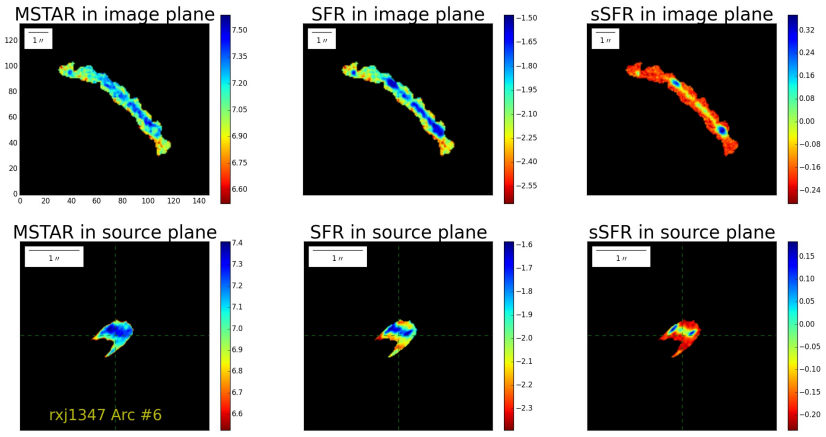
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



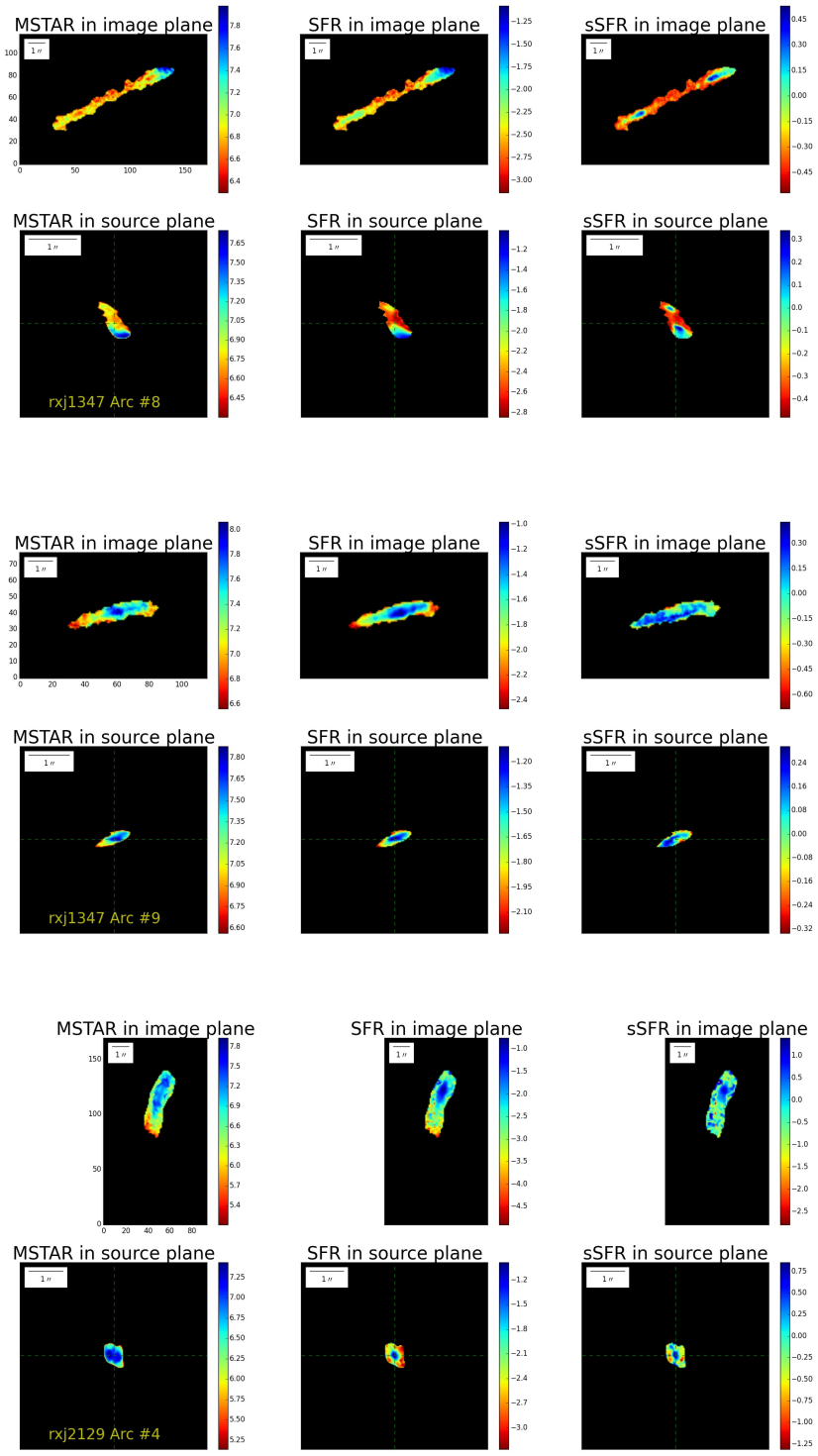
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



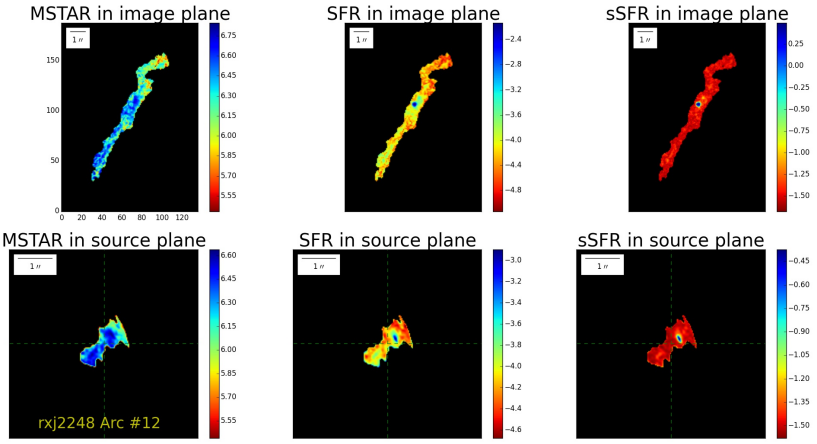


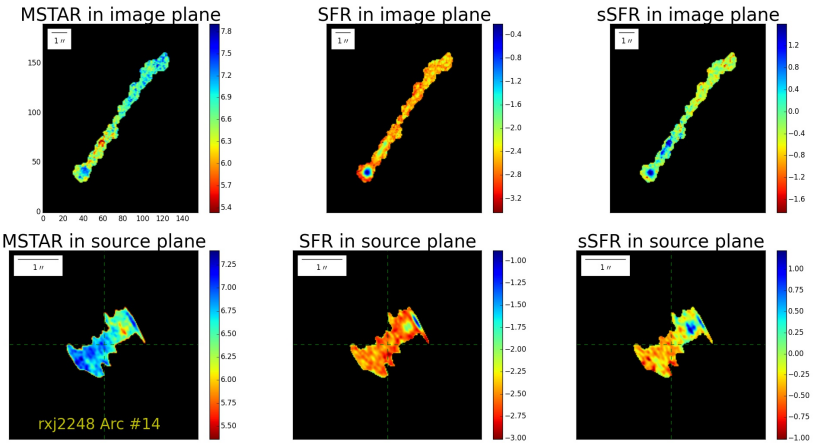


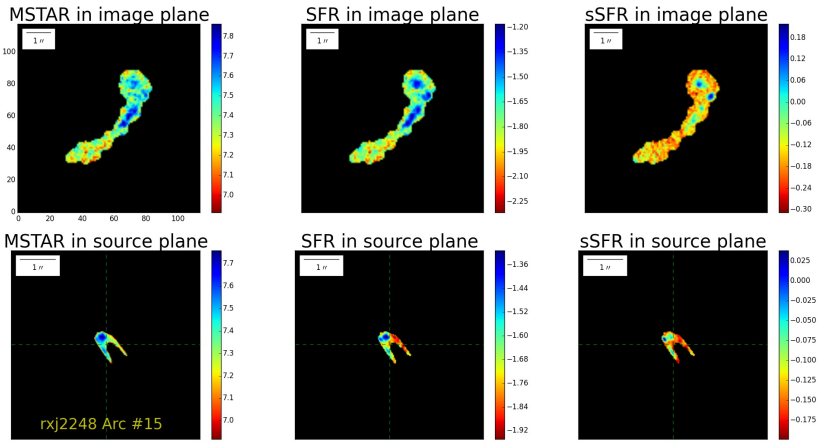
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



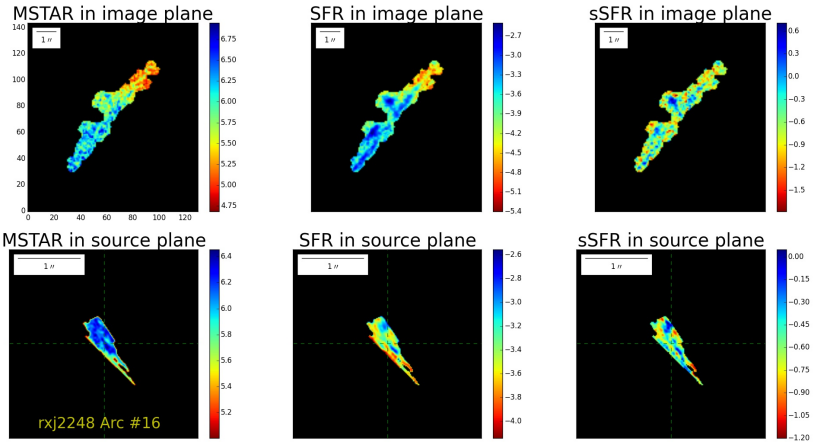
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES

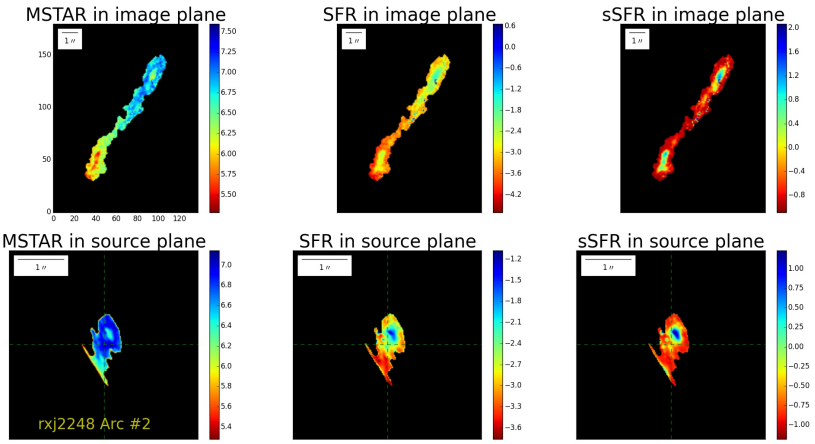


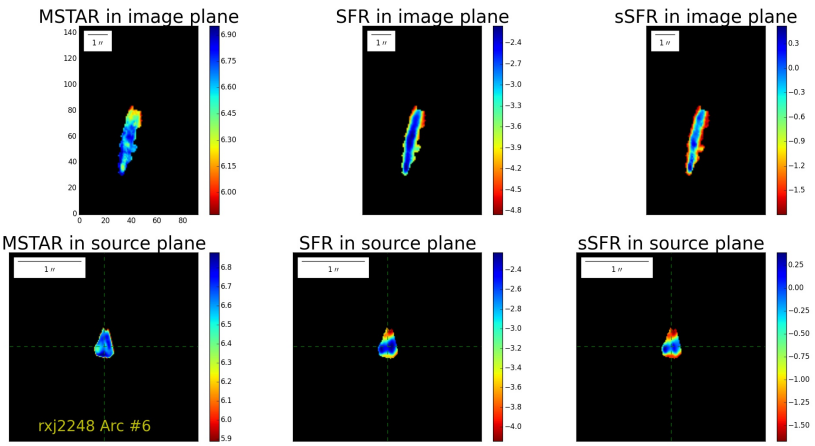




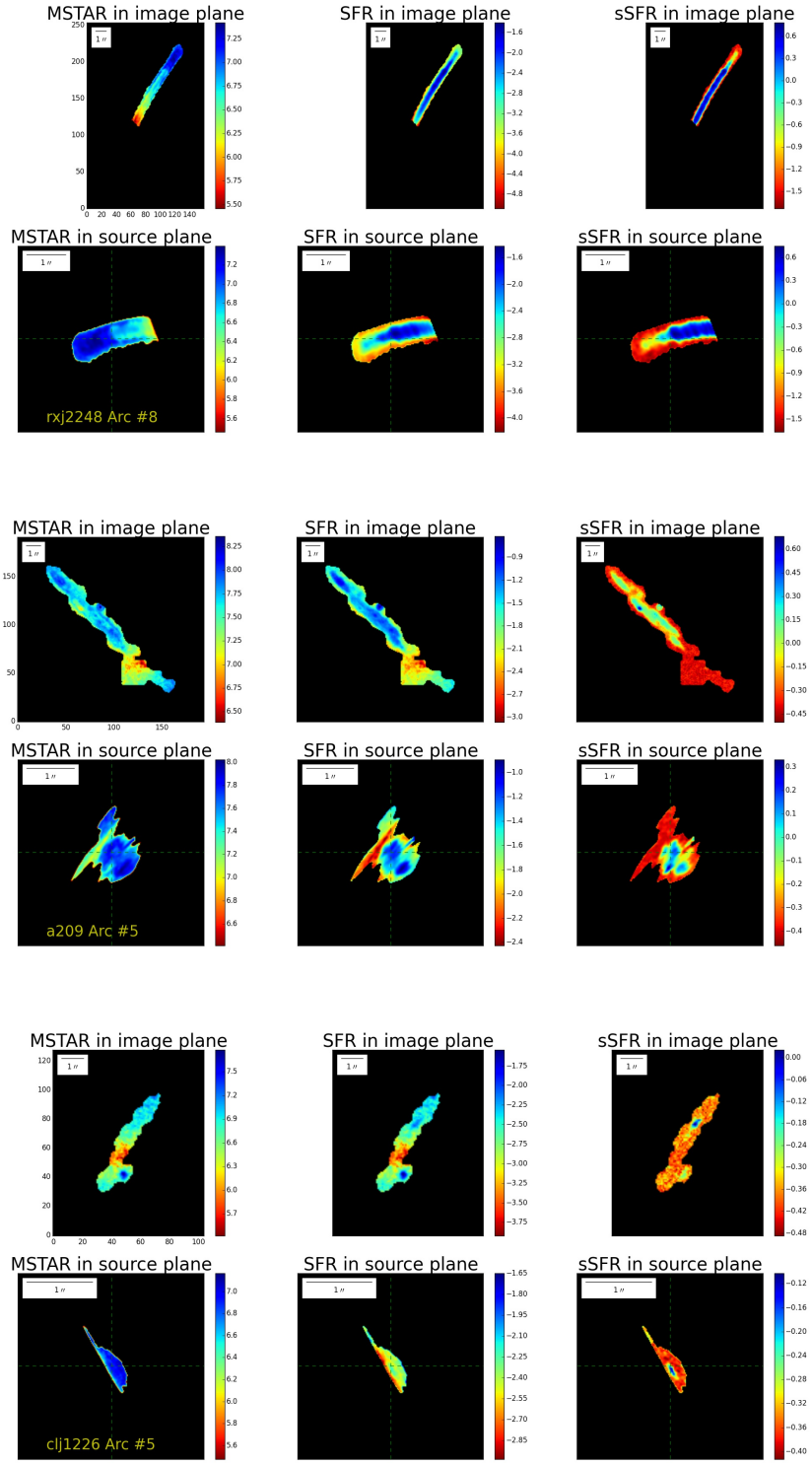
APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



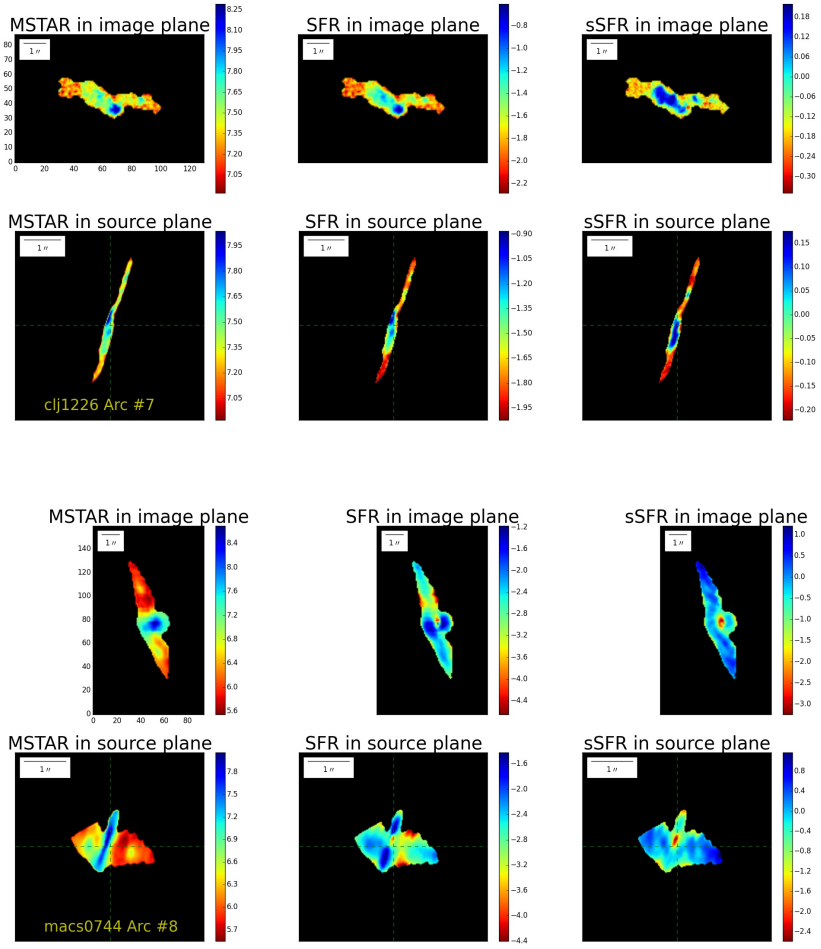




APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



APPENDIX G. THE FULL IMAGE CATALOG OF SPATIALLY RESOLVED MAPS IN IMAGE AND SOURCE PLANES



Appendix H

The Full Image Catalog of Detected Clumps

The full image catalog of detected clumps in rest-frame 2200$\mathring{A}$ along with the de-lensed images of host galaxies are listed in Figure H.1.

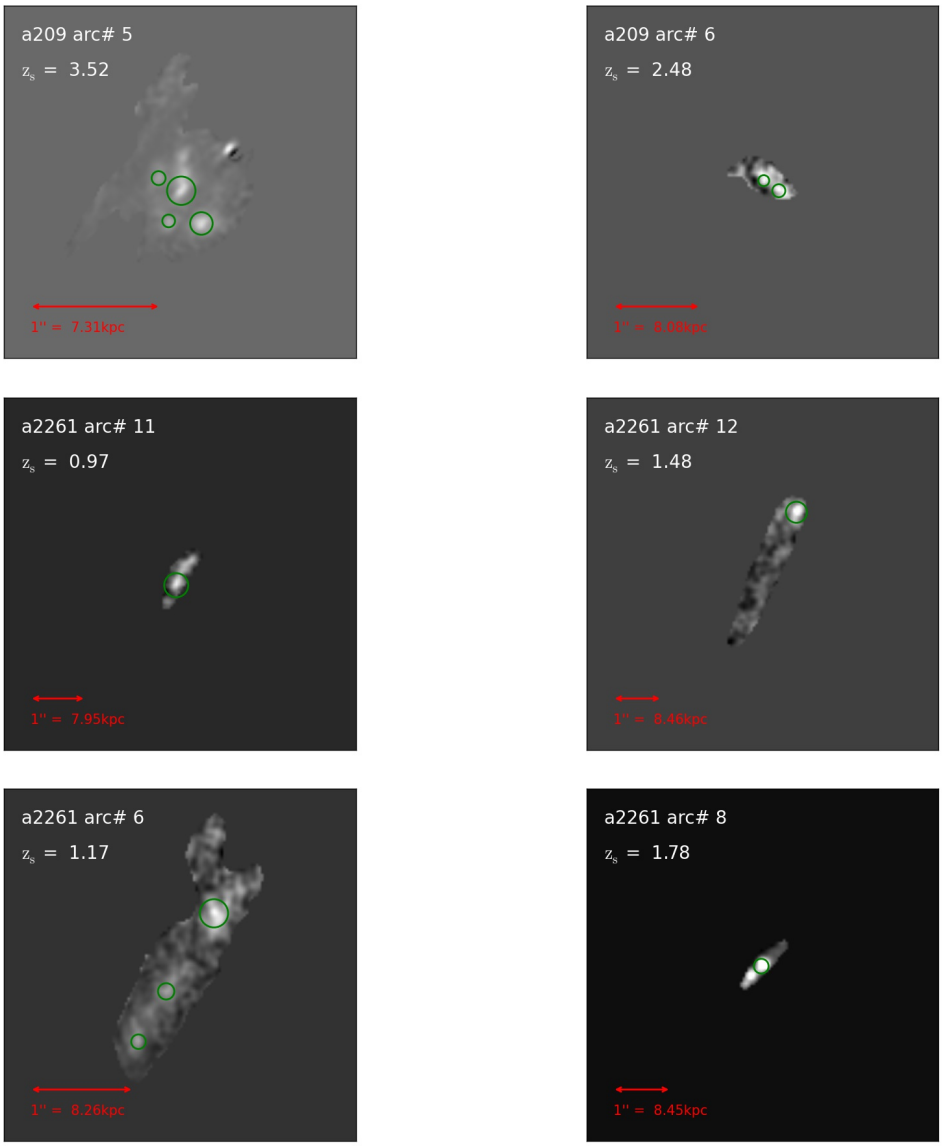
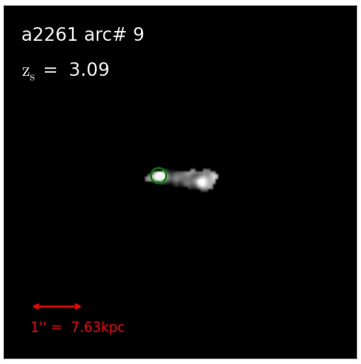
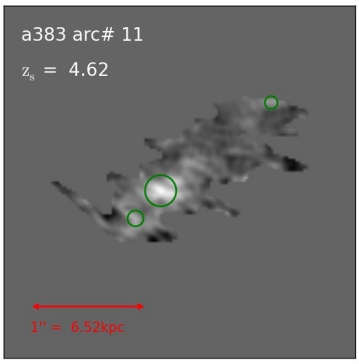
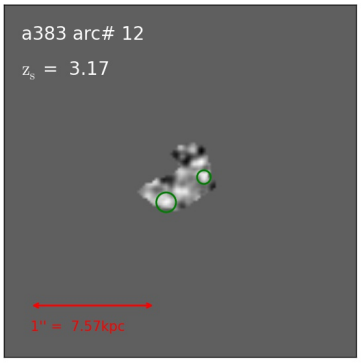


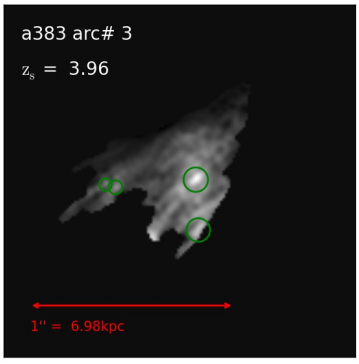
Figure H.1: The reconstructed rest-frame 2200 $\mathring{A}$ images of lensed galaxies in the source plane. The green circles mark the size and position of the detected blobs.

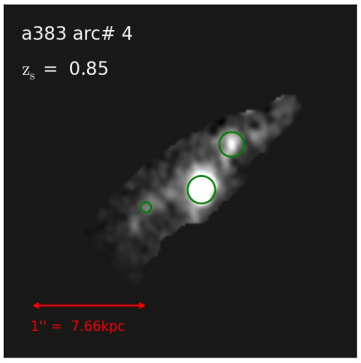
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

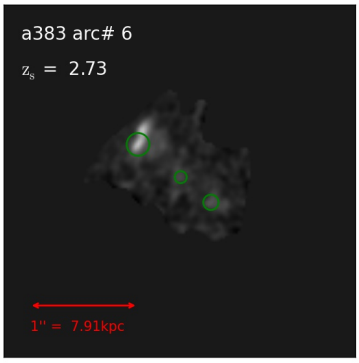




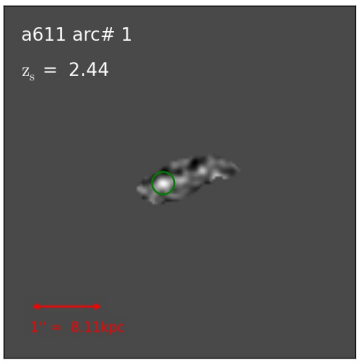


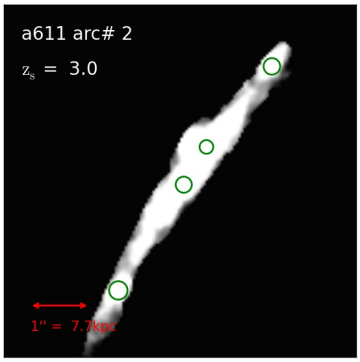


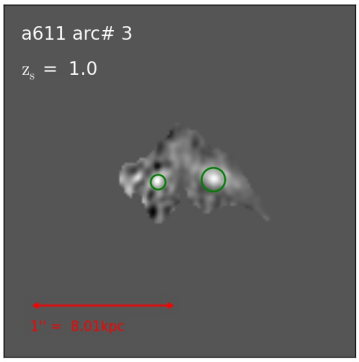


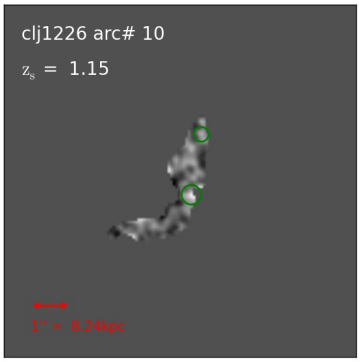


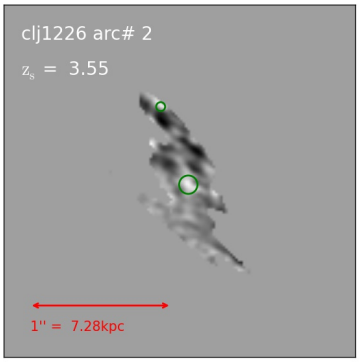
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS



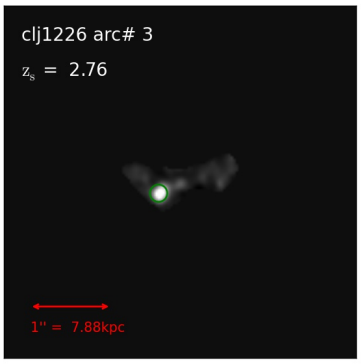


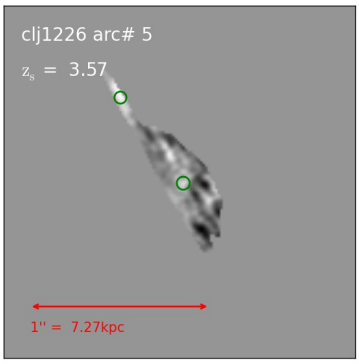


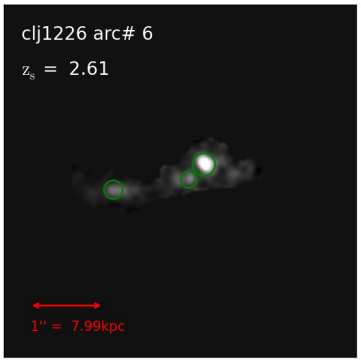


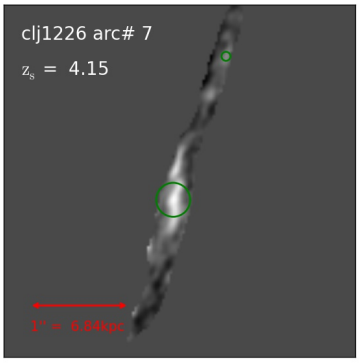


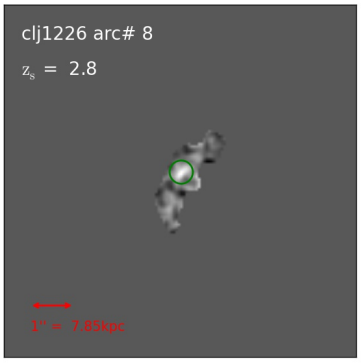
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

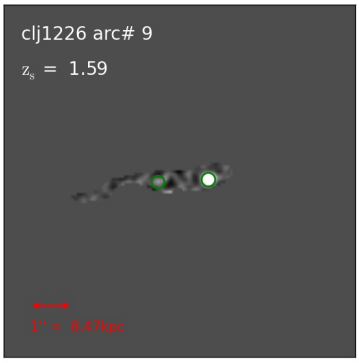




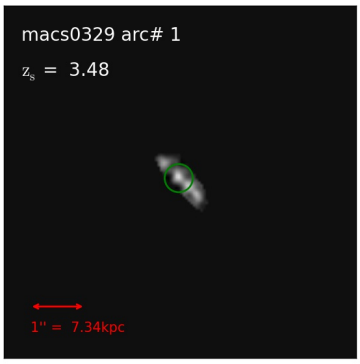


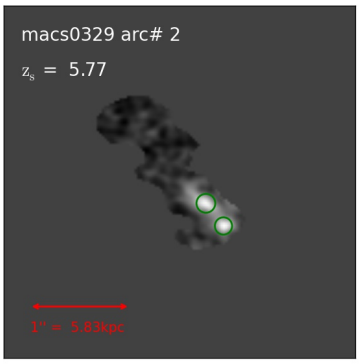


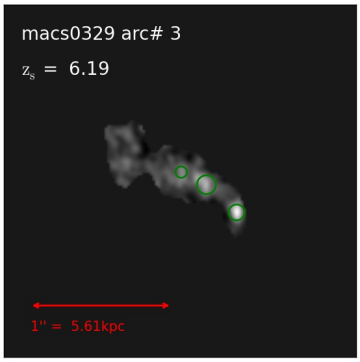


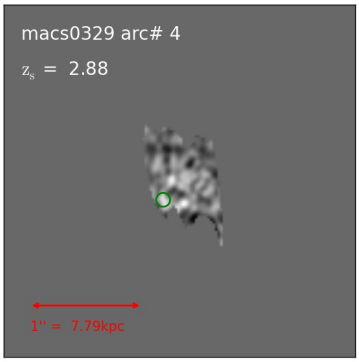


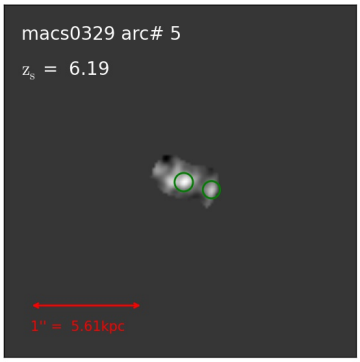
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

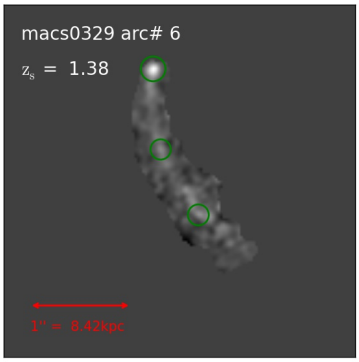




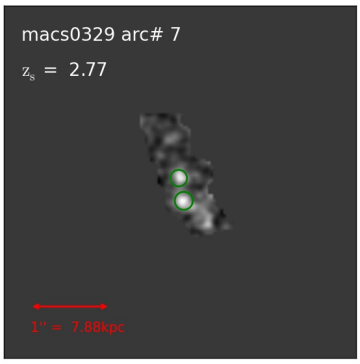


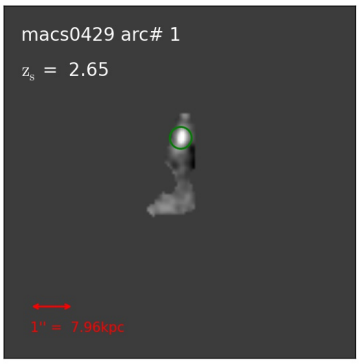


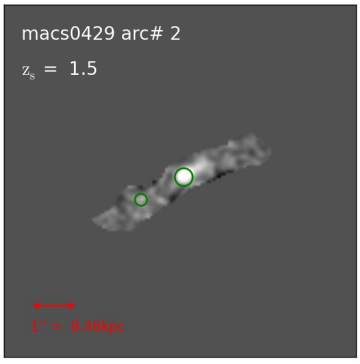


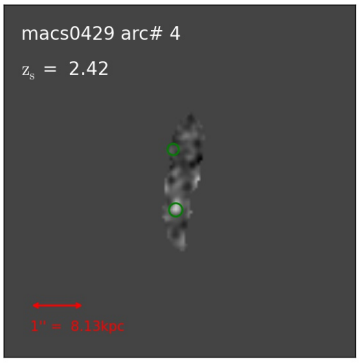


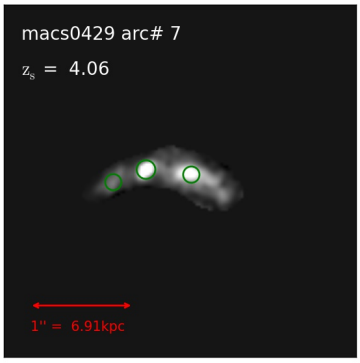
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

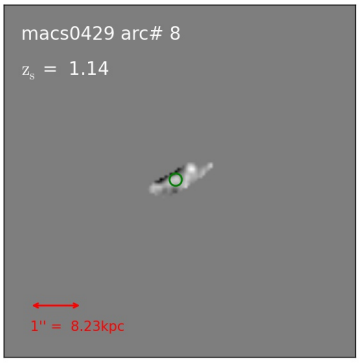




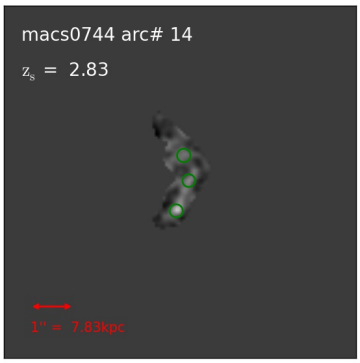


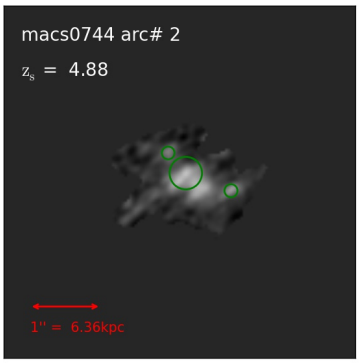


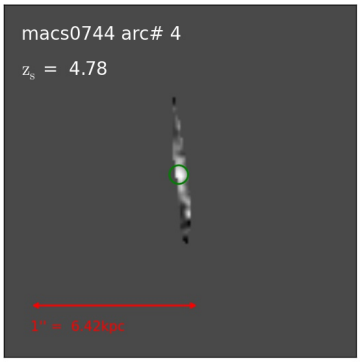


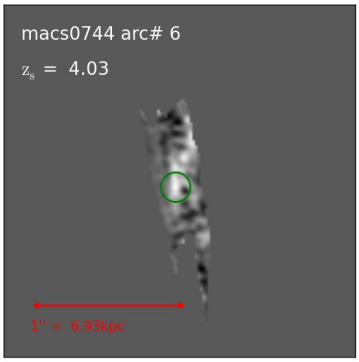


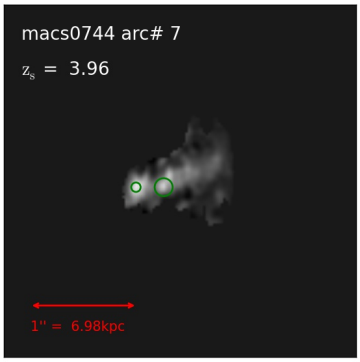
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

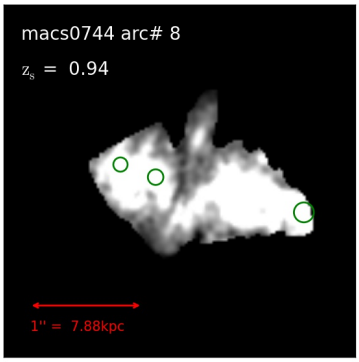




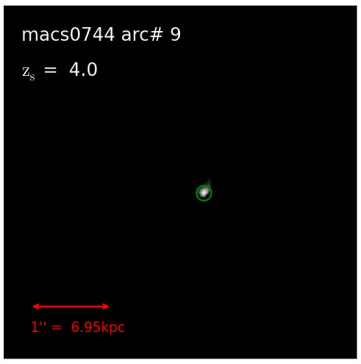


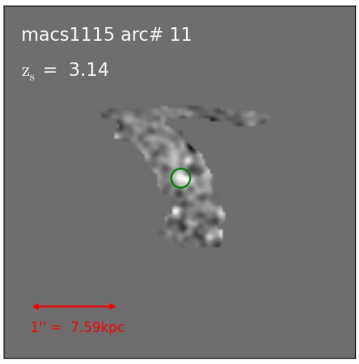


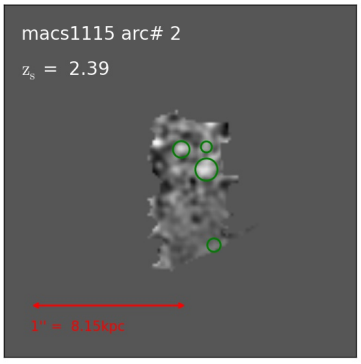


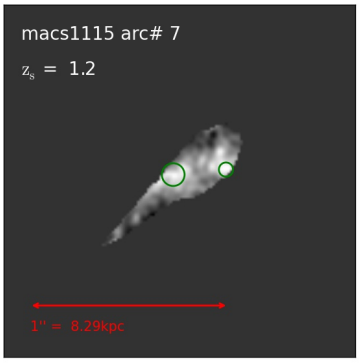


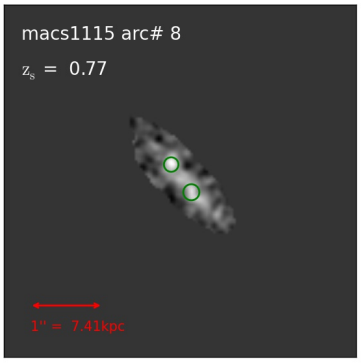
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

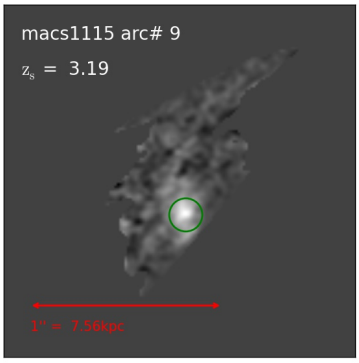




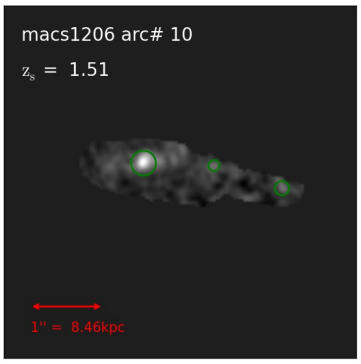


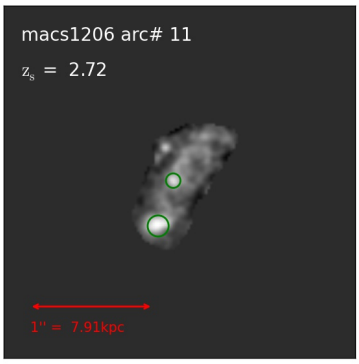


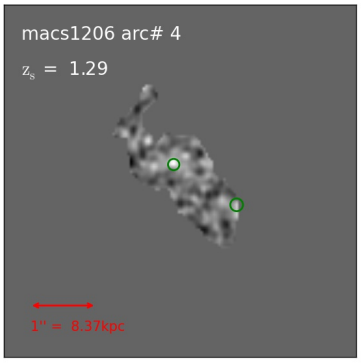


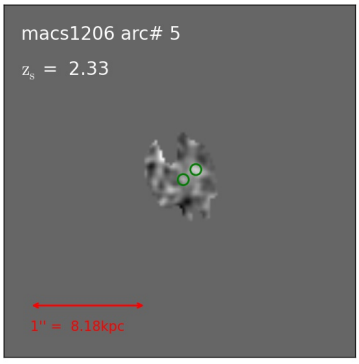


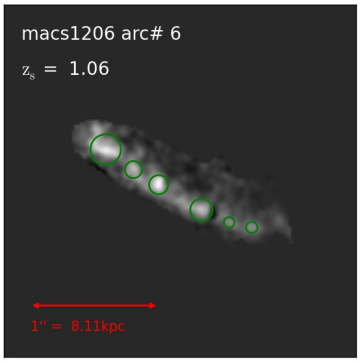
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

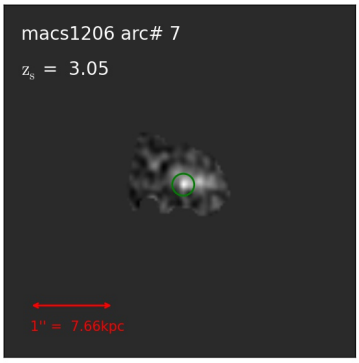




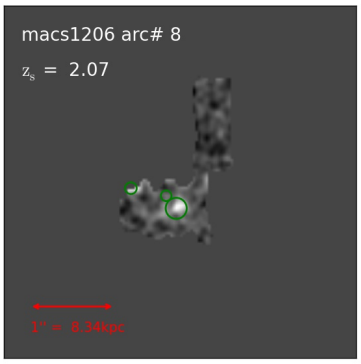


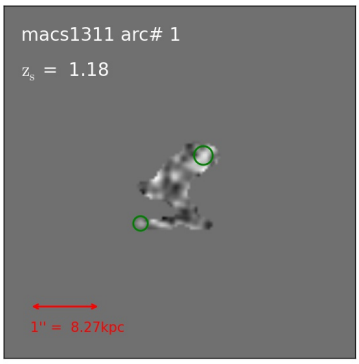


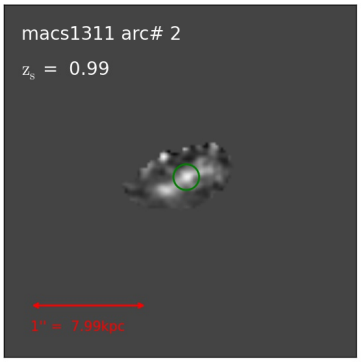


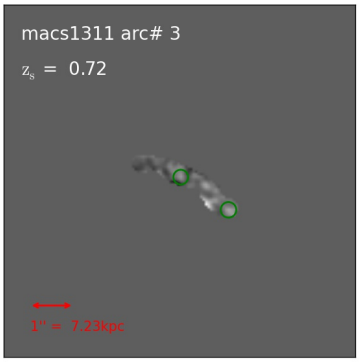


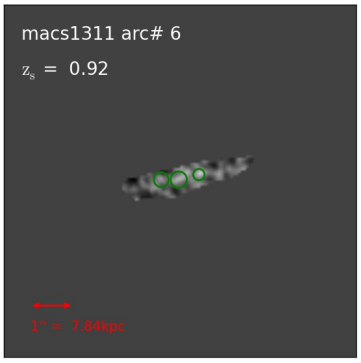
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

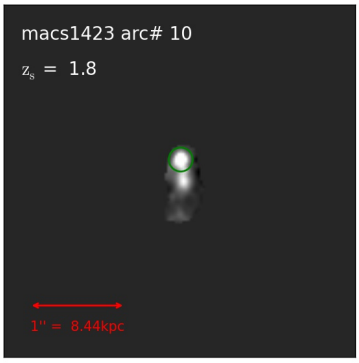




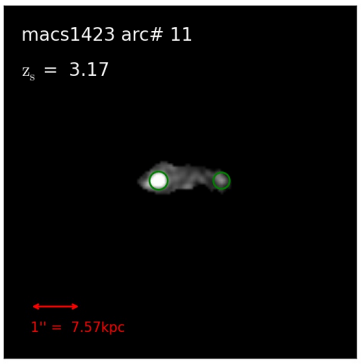


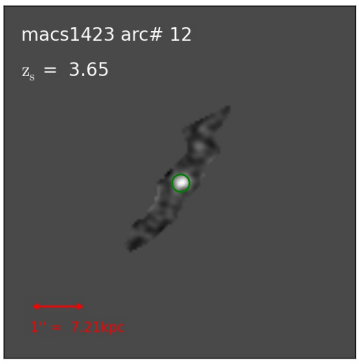


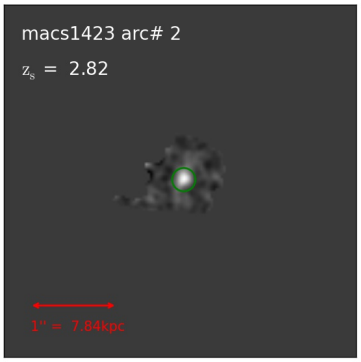


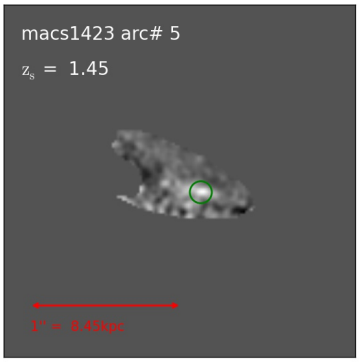


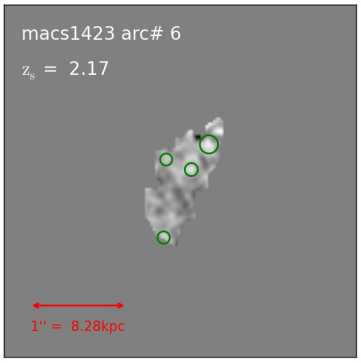
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

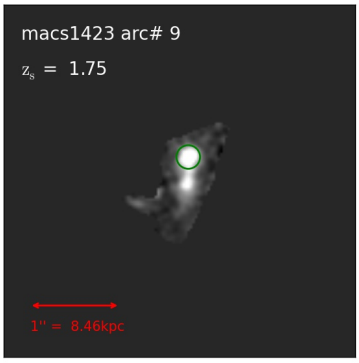




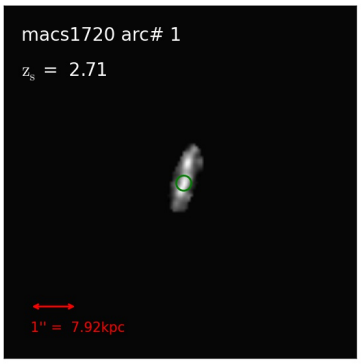


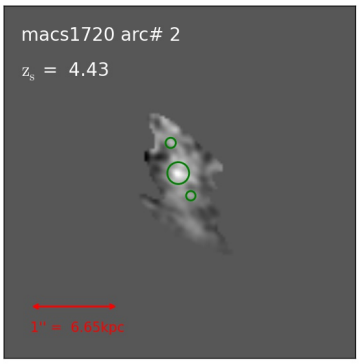


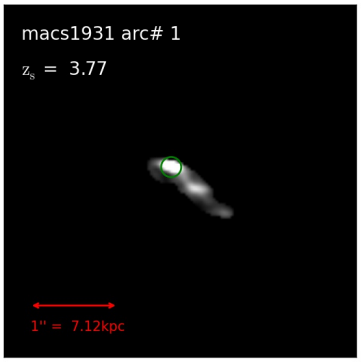


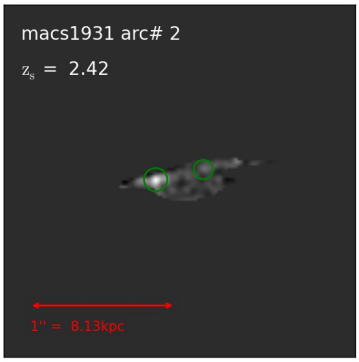


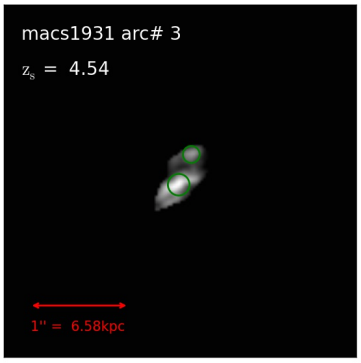
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS



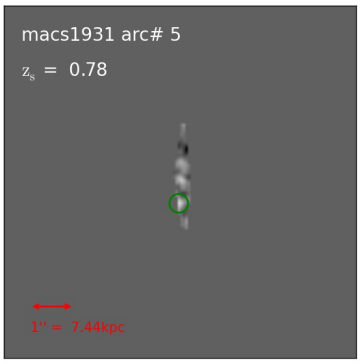


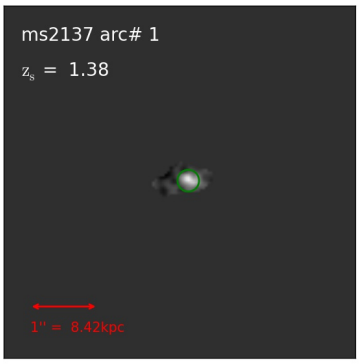


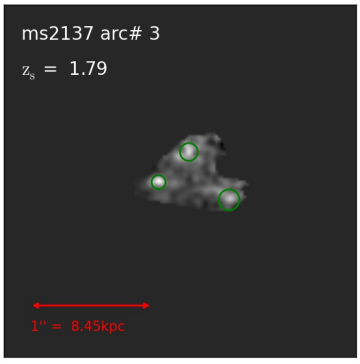


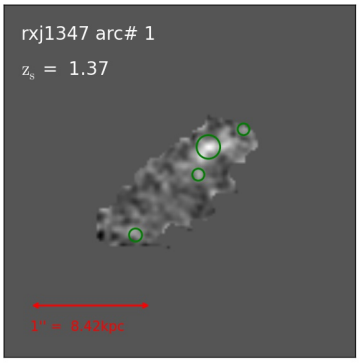


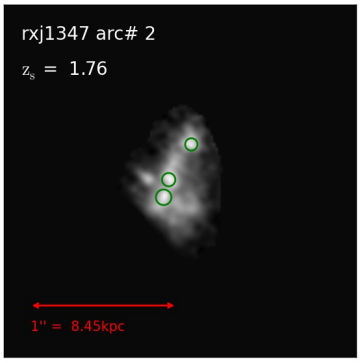
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

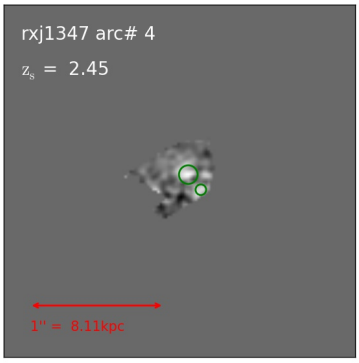




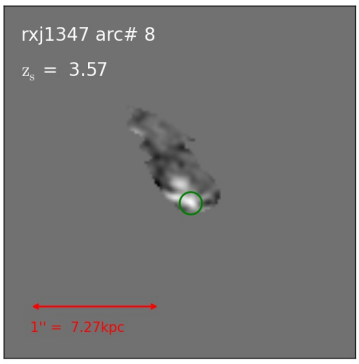


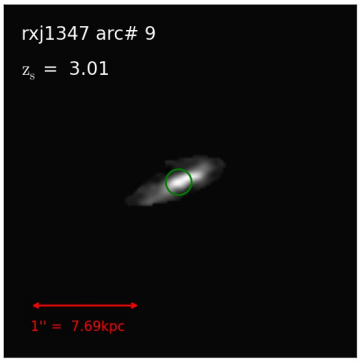


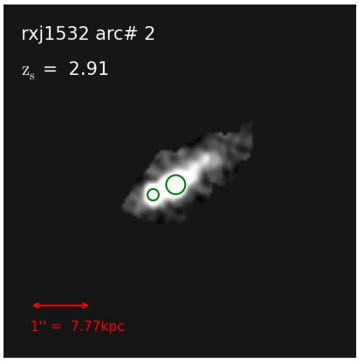


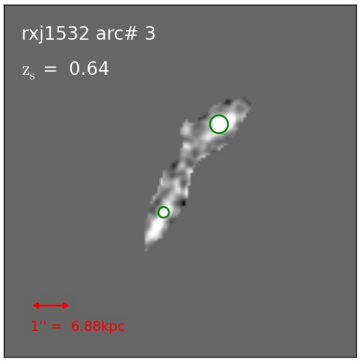


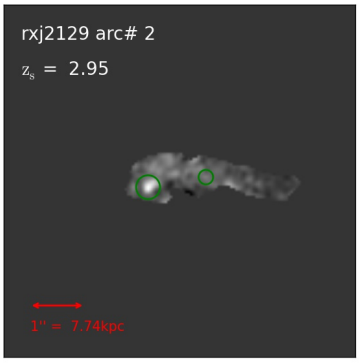
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS



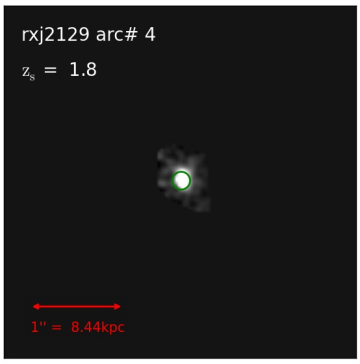


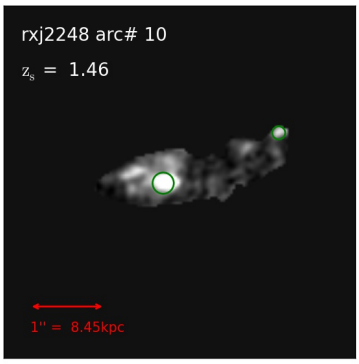


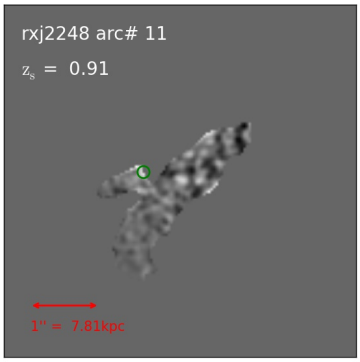


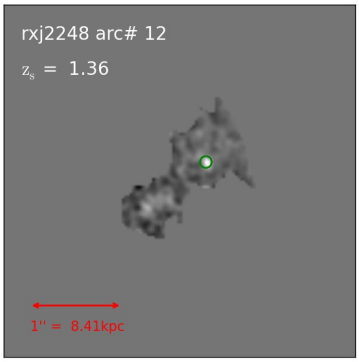


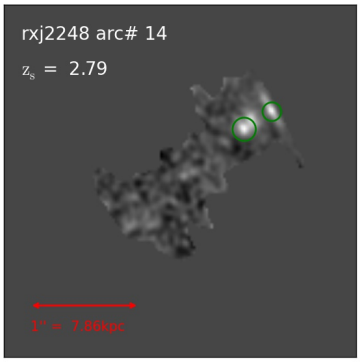
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

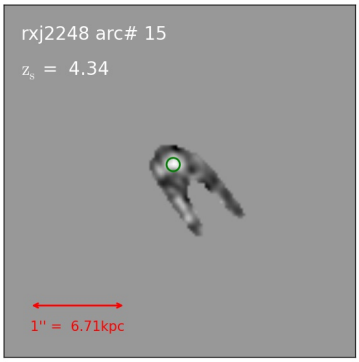




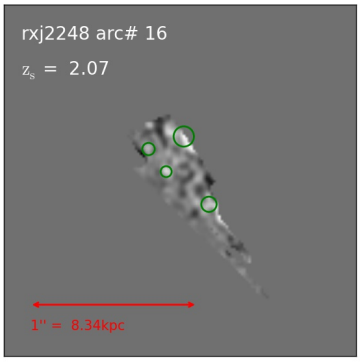


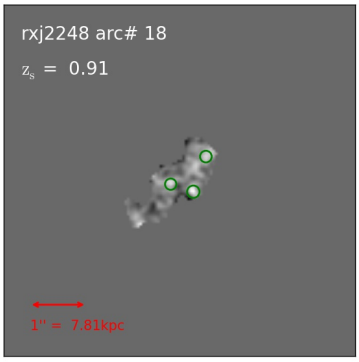


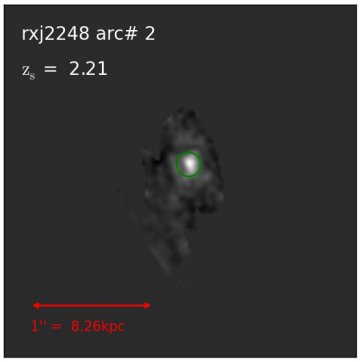


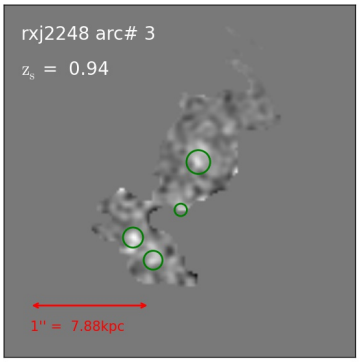


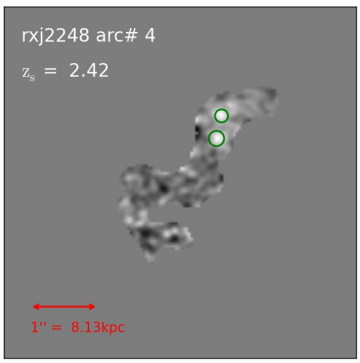
APPENDIX H. THE FULL IMAGE CATALOG OF DETECTED CLUMPS

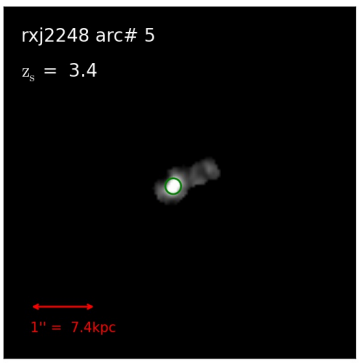


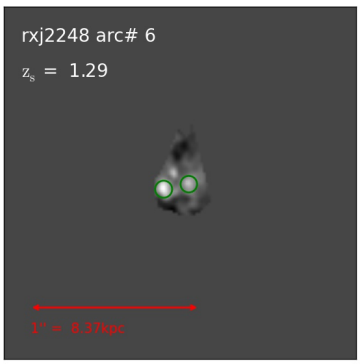


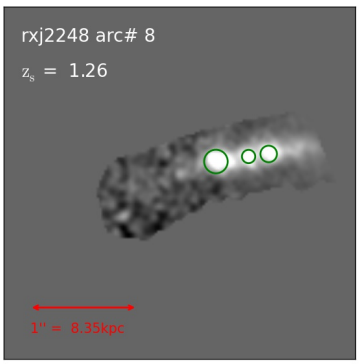












Bibliography

Abraham, R. G., Tanvir, N. R., Santiago, B. X., Ellis, R. S., Glazebrook, K., & van den Bergh, S. 1996, MNRAS, 279, L47

Abraham, R., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218

Alard, C. Arxiv e-prints (2006)

Baldry, I. K. 2008, Astronomy & Geophysic, 49, 525

Barro, G., Faber, S.M., Perez-Gonzalez, P.G. et al. 2013, ApJ, 765, 104

Barro, G., Faber, S. M., Koo, D. C., et al. 2017, ApJ, 840, 4

Bartelmann, M., & Weiss, A. 1994, A&A, 287, 1

Bartelmann, M., Huss, A., Colberg, J. M., Jenkins, A., Pearce, F. R. 1998, A&A, 330, 1

Bartelmann, M., Meneghetti, M., Perrotta, F., et al. 2003, A&A, 409, 449

Bayliss, M. B., Gladders, M. D., Oguri, M., et al. 2011, ApJ, 727, L26

BIBLIOGRAPHY

Bayliss, M. B. 2012, ApJ, 744, 156

Beasley M. A., Strader J., Brodie J. P., Cenarro A. J., Geha, M. 2006, AJ, 131, 814

Beckwith, S., Stiavelli, M., Koekemoer, A. et al. 2006, ApJ, 132, 1729

Benítez, N. 2000, ApJ, 536, 571

Benítez, N., Ford, H., Bouwens, R., et al. 2004, ApJs, 150, 1

Bhattacharya, S., Habib, S., Heitmann, K., & Vikhlinin, A. 2013, ApJ, 766, 32

Bialas, D., Lisker, T., Olczak, C., Spurzem, R., Kotulla, R. 2015, A&A, 576, A103

Biffi, V., Sembolini, F., De Petris, M., Valdarnini, R., et al. 2014, MNRAS, 439, 588

Binggeli B., Tarenghi M., Sandage A. 1990, A&A 228, 42

Binney, J.; Merrifield, M. (1998) Galactic Astronomy. Princeton: Princeton University Press. ISBN 978-0-691-02565-0.

Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349

Boada, S., Tilvi, V., Papovich, C., et al. 2015, ApJ, 803, 104

Boldrin, M., Giocoli, C., Meneghetti, M., et al. 2015 (arXiv:1505.03515)

BIBLIOGRAPHY

Bouche N., Dekel A., Genzel R., Genel S., Cresci G., Forster Schreiber N. M., Shapiro K. L., Davies R. I., Tac-coni L., 2010, ApJ, 718, 1001

Bournaud, F., Elmegreen, B. G., & Elmegreen, D. M. 2007, ApJ, 670, 237

Bournaud, F., & Elmegreen, B. G. 2009, ApJ, 694, L1

Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33

Bournaud F. et al., 2014, ApJ, 780, 57

Bouwens R. J., Illingworth G. D., Oesch P. A. et al. 2015, ApJ, 803, 34

Bowler R. A. A., Dunlop J. S., McLure R. J. et al. 2015, MNRAS, 452, 1817

Broadhurst, T. J., & Barkana, R. 2008, MNRAS, 390, 1647

Bruce, V. A. et al. 2014, MNRAS, 444, 1001

Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000

Buitrago, F., Trujillo, I., Conselice, C. J., Bouwens, R. J., Dickinson, M., Yan, H. 2008, ApJ, 687, L6

Bundy, K., Fukugita, M., Ellis, R. S., et al. 2009, ApJ, 697, 1369

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Cappellari, M. & Copin, Y. 2003, MNRAS, 342, 345

BIBLIOGRAPHY

Ceverino D., Dekel A., Bournaud F., 2010, MNRAS, 404, 2151

Ceverino, D., Dekel, A., Mandelker, N., Bournaud, F., Burkert, A., Genzel, R., Primack, J., 2012, MNRAS, 420, 3490

Ceverino D., Primack J., Dekel A., 2015, MNRAS, 453, 408

Chabrier, G. 2003, PASP, 115, 763

Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, AJ, 132, 926

Conselice, C., Bershady, M., & Jangren, A. 2000, ApJ, 529, 886

Conselice C. J., Yang C. and Bluck A. F. L. 2009, MNRAS, 394, 1956

Daddi E., et al., 2007a, ApJ 670, 15

Daddi, E., Bournaud, F., Walter, F., et al. 2010, ApJ, 713, 686

Dalal, N., Holder, G., & Hennawi, J. F. 2004. ApJ, 609, 50

D'Aloisio, A., Natarajan, P. 2011, MNRAS, 411, 1628

D'Aloisio, A., Natarajan, P. 2011, MNRAS, 415, 1913

Damjanov I., McCarthy P. J., Abraham R. G., Glazebrook, K., Yan H., Mentuch E., Le Borgne D., et al., 2009, ApJ, 695, 101

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2

Dekel, A. & Birnboim, Y. 2008, MNRAS, 383, 119

BIBLIOGRAPHY

Dekel A., Birnboim, Y., Engel, G. et al. 2009, Nature 457, 451

Dekel A., Sari R., Ceverino D., 2009b, ApJ, 703, 785 (DSC09)

de Ravel, L., Le Fevre, O., Tresse, L., et al. 2009, A & A, 498, 379

Dressler, A. & Gunn, J. E. 1983, ApJ, 270, 7

Elmegreen, D., Elmegreen, B, Rubin, D., et al. 2005, ApJ, 631, 85

Elmegreen D. M., Elmegreen B. G., Ravindranath S., Coe D. A., 2007, ApJ, 658, 763

Elmegreen, B. G., Elmegreen, D. M., Fernandez, M. X., Lemonias, J. J., et al. 2009, ApJ, 692, 12

Elmegreen, B. G., Elmegreen, D. M., Sanchez Almeida, J., Munoz-Tunon, C., Dewberry, J., Putko, J., Teich, Y., & Popinchalk, M. 2013, ApJ, 774, 8

Epinat, B., Tasca, L., Amram, P., et al. 2012, A&A, 539, A92

Erb, D. K., Shapley, A. E., Pettini, M., Steidel, C. C., et al. 2006, ApJ, 644, 813

Fedeli, C., Meneghetti, M., Bartelmann, M., Dolag, K., & Moscardini, L. 2006, A&A,447, 419

Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 326, 950

Fedeli, C., Bartelmann, M., Meneghetti, M., & Moscardini, L. 2008, A&A, 486, 35

BIBLIOGRAPHY

Finkelstein S. L., Ryan R. E. Jr., Papovich C. et al. 2015, ApJ, 810, 71

Flores, R. A., Maller, A. H., & Primack, J. R. 2000, ApJ, 535, 555

Ford, H.C. et al. 2003, Proc. SPIE, 4854, 81

Forster Schreiber, N. M., Genzel, R., Bouche, N., et al. 2009, ApJ, 706, 1364

Forster Schreiber, N. M., Shapley, A. E., Erb, D. K., Genzel, R., Steidel, C. C., Bouch, N., Cresci, G., Davies, R. 2011a, ApJ, 731, 65

Furlanetto, C., Santiago, B. X., Makler, M., et al. 2013, A&A, 549, 80

Gao, L., White, S. D. M., Jenkins, A., Stoehr, F., Springel, V. 2004, MNRAS, 355, 819

Gardini, A., Rasia, E., Mazzotta, P., Tormen, G., et al. 2004, MNRAS, 351, 505

Genel S. et al., 2012, ApJ, 745, 11

Genzel, R., Tacconi, L. J., Eisenhauer, F., Frster Schreiber, N. M., et al. 2006, Nature, 442, 786

Genzel, R. et al. 2008, ApJ, 687, 59

Genzel R., Newman S., Jones T. et al. 2011, ApJ, 733, 101

Giocoli, C., Meneghetti, M., Bartelmann, M., et al. 2012, MNRAS, 421, 3343

Gnerucci, A., Marconi, A., Cresci, G., Maiolino, R., et al. 2011, A&A, 533, 124

BIBLIOGRAPHY

Gonzlez, V., Labb, I., Bouwens, R. J., Illingworth, G., et al. 2010, ApJ, 713, 115

Gottlöber S., & Yepes G. 2007, ApJ, 664, 117

Grazian, A., et al. 2015, A&A, 575, 9

Grogin, N., Kocevski, D., Faber, S. M., Ferguson, H., et al. 2011, ApJs, 197, 35

Guo, Y., Giavalisco, M., Ferguson, H. C., Cassata, P., Koekemoer, A. M. 2012, ApJ, 757, 120

Guo, Y., Ferguson, H., Bell, E, et al. 2015, ApJ, 800, 39

Hennawi, J. F., Dalal, N., Bode, P., & Ostriker, J. P. 2007, ApJ, 654, 714

Hilbert, S., Metcalf, R. B., & White, S. D. M. 2007, MNRAS,382, 1494

Hilbert S., White S. D. M., Hartlap J., & Schneider P. 2008, MNRAS, 386, 1845

Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2008, ApJS, 175, 356

Hopkins, F. Quataert, E., Murray, N. 2012, MNRAS, 421, 3522

Horesh, A., Ofek, E. O., Maoz, D., et al. 2005, ApJ, 633, 768

Horesh, A., Maoz, D., Ebeling, H., Seidel, G., & Bartelmann, M., 2010, MNRAS, 406, 1318

BIBLIOGRAPHY

Horesh, A., Maoz, D., Hilbert, S., & Bartelmann, M., 2011, MNRAS, 418, 54

Hubble, E. P. “The realm of the nebulae”. Mrs. Hepsa Ely Silliman memorial lectures, 25. New Haven: Yale University Press. ISBN 9780300025002.

Immeli A., Samland M., Gerhard O., Westera P., 2004a, A&A, 413, 547

Immeli A., Samland M., Westera P., Gerhard O., 2004b, ApJ, 611, 20

Janz J., Laurikainen E., Laine J., Salo H., Lisker T., 2016, MNRAS, 461, L82

Jing, Y., & Suto, Y. 2002, ApJ, 574, 538

Jones T., Ellis R., Jullo E., Richard J., 2010, ApJ, 725, L176

Jouvel S., Host O., Lahav O., Seitz S., Molino A., Coe D., Postman M., Moustakas L., 2014, A&A, 562, A86

Jullo, E., Natarajan, P., Kneib, J.-P., D‘Aloisio, A., et al. Science, 329, 924

Kannappan, S. J., Guie, J. M., & Baker, A. J. 2009, ApJ, 138, 579

Keres D., Katz, N., Weinberg, D. H., Dave, R. 2005, MNRAS 363, 2

Killedar, M., Borgani, S., Meneghetti, M., et al. 2012, MNRAS, 427, 533

Kimble, R. A. et al. 2008, Proc. SPIE, 7010, 1

Klypin, A., Kravtsov, A. V., Bullock, J. S., & Primack, J. R. 2001, ApJ, 554, 903

Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, ApJs, 111, 73

BIBLIOGRAPHY

Koekemoer, A., Faber, S. M., Ferguson, H., Grogin, N., et al. 2011. ApJs, 197, 36

Komatsu, E.; Smith, K. M.; Dunkley, J.; et al. 2011, ApJS, 192, 18

Kormendy, J., Fisher, D. B., Cornell, M. E., Bender, R. 2009, ApJS, 182, 216

Kormendy, J. 1985, ApJ, 295, 73

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692

Larson, R. B. 1981, MNRAS, 194, 809

Lenzen, F., Schindler, S., & Scherzer, O. 2004, A&A, 416, 391

Li, G.-L., Mao, S., Jing, Y. P., et al. 2005, ApJ, 635, 795

Li, G.-L., Mao, S., Jing, Y. P., et al. 2006, MNRAS 372, 73

Livermore R. C. et al., 2012a, MNRAS, 427, 688

Livermore R. C. et al. 2015, MNRAS, 450, 1812

Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113

Lisker, T., Glatt, K., Westera, P., & Grebel, E. K. 2006, AJ, 132, 2432

Lorenz, M. O. 1905, J. Am. Stat. Assoc., 9, 209

Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163

BIBLIOGRAPHY

Lotz J. M., Jonsson P., Cox T. J. & Primack J. R. 2008, MNRAS, 391, 1137

Lotz J. M., Jonsson P., Cox T. J. et al. 2011, ApJ, 742, 103

Madau, P., Ferguson, Henry C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388

Madau, P. & Dickinson, M. 2014, ARA&A, 52, 415

Mandelker N., Dekel A., Ceverino D., Tweed D., Moody C. E., Primack J., 2014, MNRAS, 443, 3675

Mandelker N., Dekel A., Ceverino D., et al. 2017, MNRAS, 464, 635

Meneghetti, M., Bolzonella, M., Bartelmann, M., Moscardini, L., & Tormen, G. 2000, MNRAS, 314, 338

Meneghetti, M., Bartelmann, M., & Moscardini, L. 2003, MNRAS, 346, 67

Meneghetti, M., Bartelmann, M., Dolag, K., Moscardini, L., et al. 2005, A&A, 442, 413

Meneghetti, M., Argazzi, R., Pace, F., Moscardini, L., et al. 2007, A&A, 461, 25

Meneghetti, M., Melchior, P., Grazian, A., De Lucia, G., et al. 2008, A&A, 482, 403

Meneghetti, M., Fedeli, C., Pace, F., Gottlóber, S., & Yepes, G. 2010, A&A, 519, 90

BIBLIOGRAPHY

Meneghetti, M., Fedeli, C., Zitrin, A., Bartelmann, M., Broadhurst, T., Gott'BFber, S., Moscar-dini, L., Yepes, G. 2011, A&A 530, 17

Meneghetti, M., Rasia, E., Vega, J., Merten, J., Postman, M., et al. 2014, ApJ, 797, 34

Merten, J., Meneghetti, M., Postman, M. et al. 2015, ApJ, 806, 4

Milos, J., & Hernquist, L. 1996, ApJ, 464, 641

Miralda-Escude, J. 1993, ApJ, 403, 497

Misgeld, I., & Hilker, M. 2011, MNRAS, 414, 3699

Moffett A. J., Lange R., Driver S. P. et al. 2016, MNRAS, 462, 4336

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, Nature, 379, 613

Moustakas, J., Coil, A. L., Aird, J., Blanton, M. R., Cool, R. J., et al. ApJ, 2013, 767, 50

Murray N., Quataert E., Thompson T. A., 2010, ApJ, 709, 191

Narayan, R. and Bartelmann, M. (1996) "Lectures on Gravitational Lensing". astro-ph/9606001

Nesvadba N. P. H. et al., 2006, ApJ, 650, 661

BIBLIOGRAPHY

Neto, A. F., Gao, L., Bett, P., et al. 2007, MNRAS, 381, 1450

Noeske K. G., et al., 2007, ApJ, 660, L4

Noguchi M., 1999, ApJ, 514, 77

Oguri, M., Lee, J., & Suto, Y. 2003, ApJ, 559, 572

Oguri, M., & Blandford, R. 2009, MNRAS, 392, 930

Oklopcic A., Hopkins P. F., Feldmann R., Keres D., et al. 2017, 465, 952

Pan Z., Li J., Lin W. et al 2015, ApJL, 804, L42

Papovich C., et al., 2015, ApJ, 803, 26

Peng, C. Y., Ho, L. C., Impey, C. D., Rix, H. 2002, AJ, 124, 266

Planck Collaboration, Ade, P. A, Aghanim, N., Armitage-Caplan, C., et al. 2014, A&A, 571, A16

Postman, M., Coe, D., Benitez, N., et al. 2012, ApJs, 199, 25

Puchwein, E., Bartelmann, M., Dolag, K., & Meneghetti, M. 2005, A&A, 442, 405

Puchwein, E., & Hilbert, S. 2009, MNRAS,398, 1298

Redlich, M., Bartelmann, M., Waizmann, J.-C., Fedeli, C. 2012, A&A 547, 66

Renzini, A. 2009, MNRAS, 398, L58

BIBLIOGRAPHY

Richard, J., Smith, G. P., Kneib, J. et al. 2010, MNRAS, 404, 325

Rigby J. R., Johnson T. L., Sharon K. et al. 2017, ApJ, 843, 79

Rubin, Vera C.; Ford, W. Kent, Jr. 1970, ApJ, 159, 379

Robaina A. R., Bell E. F., Skelton R. E., McIntosh D. H., Somerville R. S., Zheng X., Rix H.-W., Bacon D., Balogh M., Barazza F. D., Barden e. a., 2009, ApJ, 704, 324

Rodighiero G., Daddi E., Baronchelli I., Cimatti A., Renzini A., Aussel H., Popesso P., Lutz D., Andreani P., Berta S., Cava A., Elbaz D., Feltre A., Fontana A., Forster Schreiber e. a., 2011, ApJ, 739, L40

Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, ApJ, 723, 492

Rozo, E., Nagai, D., Keeton, C., & Kravtsov, A. 2008, ApJ, 687, 22

Saintonge A., Kauffmann G., Kramer C. et al. 2011, MNRAS, 415, 32

Sandage, A. 1986, A&A, 161, 89

Santini P., Ferguson H. C., Fontana A., et al. 2015, ApJ, 801, 97

Seidel, G., & Bartelmann, M. 2007, A&A, 472, 341

Sembolini, F., Yepes, G., De Petris, M., Gottlober, S., et al. 2013 MNRAS, 429, 323

BIBLIOGRAPHY

Sereno, M., Jetzer, P., & Lubini, M. 2010, MNRAS, 403, 2077

Schneider, P. 1992 A & A, 260, 1

Schechter, P. 1976, ApJ, 203, 297

Shapiro, K. L., et al. 2008, ApJ, 682, 23

Shibuya, T., Ouchi, M., Kubo, M., Harikane, Y. 2016, ApJ, 821, 72

Smith, R. J., Lucey, J. R., Hudson, M. J., et al. 2009, MNRAS, 392, 1265

Somerville, S., Lee, K., Ferguson, H. et al. 2004, ApJ, 600, L171

Soto, E., de Mello, D. F., Rafelski, M., et al. 2017, ApJ, 837, 6

Springel, V. 2005, MNRAS, 364, 1105

Stark, D. P., Ellis, R. S., Bunker, A., Bundy, K., Targett, T., Benson, A., & Lacy, M. 2009, ApJ, 697, 1493

Swinbank A. M. et al., 2009, MNRAS, 400, 1121

Swinbank, A. M., Smail, I., Longmore, S., Harris, A. I., et al. 2010, Nature, 464 733

Swinbank A. M., Sobral D., Smail I., Geach J. E., Best P. N., McCarthy, I. G., Crain R. A., Theuns T., 2012, MNRAS, 426, 935

Szomoru, D., Franx, M., Bouwens, R., et al. 2011, ApJ, 735, L22

BIBLIOGRAPHY

Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246

Tacconi, L. J., Neri, R., Genzel, R., Combes, F., Bolatto, A., et al. 2013, ApJ, 768, 74

Toomre, A. 1964, ApJ, 139, 1217

Toomre, A., & Toomre, J. 1972, ApJ, 178, 623

Torri, E., Meneghetti, M., Bartelmann, M., et al. 2004, MNRAS, 349, 476

Trujillo, I., Conselice, C. J., Bundy, K., et al. 2007, MNRAS, 382, 109

Tully R. B. & Trentham N. 2008, AJ, 135, 1488

Umetsu, K., Medezinski, E., Nonino, M., Mertem, J., Postman, M., et al. 2014,ApJ, 795, 163

Vanzella, E., Calura, F., Meneghetti, M., et al. 2017, MNRAS, 467, 4304

Vogelsberger M. et al., 2014a, Nature, 509, 177

Wambsganss, J., Bode, P., & Ostriker, J. P. 2004, ApJ, 606, L93

Wambsganss, J., Bode, P., & Ostriker, J. P. 2005, ApJ, 635, L1

Wang T., Huang J. S., Faber S. M. et al. 2012, ApJ, 752, 134

Williams R. J., Quadri R. F. & Franx M. 2011, ApJL, 738, L25

Windhorst, R. A., et al. 1994c, AJ, 107, 930

BIBLIOGRAPHY

Wisnioski, E., Glazebrook, K., Blake, C., et al. 2012, MNRAS, 422, 3339

Wuyts, S., Labbé , I., Schreiber, N. M. F., et al. 2008, ApJ, 682, 985

Wuyts S., Forster Schreiber N. M., van der Wel A., et al. 2011, ApJ, 742, 96

Wuyts S., Forster Schreiber, N. M., Genzel, R., et al. 2012, ApJ, 753, 114

Wuyts, E., Rigby, J. R., Gladders, M. D., & Sharon, K. 2014, ApJ 781, 61

Zamojski, M. A., Schiminovich, D., Rich, R. M., Mobasher, B., Koekemoer, A. M. et al. 2007, ApJS, 172, 468

Zitrin, A., Meneghetti, M., Umetsu, K., Broadhurst, T., et al. 2013, ApJ, 762, 30

Zitrin, A., Fabris, A., Merten, J., et al. 2015, ApJ, 801, 44

Zolotov, A., Dekel, A., Mandelker, N., Tweed, D., et al. 2015, MNRAS, 450, 2327